

Pred. No. is the number of results predicted by chance to have a

ORGANISM

Rickettsia cells

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiinae; Rickettsia; spotted fever group. 1 (bases 1 to 5513)

REFERENCE

AUTHORS

Bouyer, D.H., Stenos, J., Crocquet-Valdes, P., Moron, C.G., Popov, Zavala-Velazquez, J.E., Foil, L.D., Stothard, D.R., Azad, A.F. and

Pred. No. is the number of results predicted by chance to have a

Walker, D.H.
 Rickettsia felis: molecular characterization of a new member of the
 spotted fever group
 Int. J. Syst. Evol. Microbiol. 51 (Pt 2), 339-347 (2001).
 MEDLINE
 21217364
 PUBMED
 11321078
 2 (bases 1 to 5513)
 Bouyer, D.H., Stenos, J., Crocquet-Valdes, P.A., Foil, L.D. and
 Walker, D.H.
 Direct Submission
 Submitted (30-SEP-1999), Pathology, University of Texas Medical
 Branch at Galveston, 301 University Blvd., Galveston, TX
 77555-0609, USA

FEATURES

source
 1. 5513
 /organism="Rickettsia felis"
 /db_xref="taxon:42862"
 478. 2337
 /gene="ompA"
 478. 2337
 /gene="ompA"
 /note="truncated as compared to other Rickettsia ompA
 proteins"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein A"
 /protein_id="AAG17185.1"
 /db_xref="GI:10441769"
 /translation="MANISLKFQKAIQKLTALFTTSTAIAIMLGGVIGAAVTG
 ADCAELAAAGTNGICPGAGFAGVAGTLOYTGAFVTDADVSRLDLDINFAGLFVTD
 ISLGSVDVGGKLANVDDGLTLTGTAAGYANPALFQGGQAANNRYETALIG
 NYTLGGANAGLTASDPVLGPIFLAGNDGGIITDNDAAINGTNGTNGTNGTNGT
 GASTLSGGAVIKATTKLTNAAPVLTNANAVLTCAVDNVTGGDDVGVNLNGALS
 QVTGNTGNTSLATISVAGTATLGAVIKATTKLTNAASVLTNAVLTGAVDNT
 GDNVGVNLGALSQVGTNGTNSLATINIGAGTLDGAVIKATTKLTDDASVL
 IFNPVVTGAIDONTNANKGVVFTGASTVDNIGNTAVLAEVSVAGLQIQGVV
 KANALNTONASVVTGDSVTGSGIGTFLPATVIGITLRAGSLAANNIDFGA
 ASNLNRPAGKNYNLIGTIANGLNATNAAAGTVIANDVSIGTVQAQINQNKIFV
 INKANADVDLDQAQISFKAARSLFANVSLQMIELSLKLIIPVLLTVVVS"
 BASE COUNT 1814 a 961 c 1181 g 1557 t
 ORIGIN

Query Match 100.0%; Score 1920; DB 1; Length 5513;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCGGAATATTTCTTAAATATTTCAAAAAGCAATTCAAAAGGTCTTAAACACTGCT 60
 DB 478 ATGCGGAATATTTCTTAAATATTTCAAAAAGCAATTCAAAAGGTCTTAAACACTGCT 537
 QY 61 TTATTCACCACTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCTTGGTGGTGA 120
 DB 538 TTATTCACCACTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCTTGGTGGTGA 597
 QY 121 AGAACCTGAAGTCTGATGTCAGAGCTTGCACCGCGAACAATATAGGTCCTGGAGCC 180
 DB 598 AGAACCTGAAGTCTGATGTCAGAGCTTGCACCGCGAACAATATAGGTCCTGGAGCC 657
 QY 181 GGTGCTTTTGTAGCGGGTCTTACATATATACCAATATACCGGTGCTTTACGGTTACTGATGCT 240
 DB 658 GGTGCTTTTGTAGCGGGTCTTACATATATACCAATATACCGGTGCTTTACGGTTACTGATGCT 717
 QY 241 GACGTAAAGTGTCTGATGATTTAAATTAATTTTCAGCAGGTCTTTTTCAGTAACT 300
 DB 718 GACGTAAAGTGTCTGATGATTTAAATTAATTTTCAGCAGGTCTTTTTCAGTAACT 777
 QY 301 GGTGATATTCATAGGTCTGATGATACCGGAGAGCTATATAACTTGCAGTTAAT 360
 DB 778 GGTGATATTCATAGGTCTGATGATACCGGAGAGCTATATAACTTGCAGTTAAT 837
 QY 361 ATTGATGATGTTTAACTTAACTTTAACAGGTACCGGTACTGCGAGCTACGGTGCCTTAACT 420
 DB 838 ATTGATGATGTTTAACTTAACTTTAACAGGTACCGGTACTGCGAGCTACGGTGCCTTAACT 897

QY 421 CCTCCGTTGTTATTCAGGTGGACAAGCTGCTCTAATAATACATATACCTGCTTTAGT 480
 DB 898 CCTCCGTTGTTATTCAGGTGGACAAGCTGCTCTAATAATACATATACCTGCTTTAGT 957
 QY 481 AATATACTCTAGTGGAGGCAATGCGGTTTGACTATTGCTTCAGATCCAGATGATTA 540
 DB 958 AATATACTCTAGTGGAGGCAATGCGGTTTGACTATTGCTTCAGATCCAGATGATTA 1017
 QY 541 GGACCAATAACGCTTCAGGAAATATAGATGGAGGATATAAATACCTGACAAATACAGAT 600
 DB 1018 GGACCAATAACGCTTCAGGAAATATAGATGGAGGATATAAATACCTGACAAATACAGAT 1077
 QY 601 GCTGCCATTAAACGGAACAATAGTAACTAATCCGCGAGCTCAAAATAGCAATTTGGAGCA 660
 DB 1078 GCTGCCATTAAACGGAACAATAGTAACTAATCCGCGAGCTCAAAATAGCAATTTGGAGCA 1137
 QY 661 AGTACGCTTCTCTTGGAGGGCAGTTATTAAAGCCACTAGGACTAAGTTAAACAATGCA 720
 DB 1138 AGTACGCTTCTCTTGGAGGGCAGTTATTAAAGCCACTAGGACTAAGTTAAACAATGCA 1197
 QY 721 GCGCGGTATTAACACTTTACAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
 DB 1198 GCGCGGTATTAACACTTTACAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1257
 QY 781 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 1258 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
 QY 841 AATATAGTAACTAATAATTCATTAGCCACCAATAGTGTAGGAGCAGGTACCGGTACGTTA 900
 DB 1318 AATATAGTAACTAATAATTCATTAGCCACCAATAGTGTAGGAGCAGGTACCGGTACGTTA 1377
 QY 901 GGGGAGCGGTATTAAAGCCACTACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 960
 DB 1378 GGGGAGCGGTATTAAAGCCACTACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1437
 QY 961 CTTACAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1020
 DB 1438 CTTACAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1497
 QY 1021 GTCGTAATCTAAGTGGAGCAATTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1080
 DB 1498 GTCGTAATCTAAGTGGAGCAATTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1557
 QY 1081 TTAGCCACGATAATATAGGAGCAGGTGCTGCTACCTTGGATGGAGCGGTATTAAAGCT 1140
 DB 1558 TTAGCCACGATAATATAGGAGCAGGTGCTGCTACCTTGGATGGAGCGGTATTAAAGCT 1617
 QY 1141 ACTACGACTAAGTTAACAGATGATGCTGCTAGTATTAATTTACGAATCCTGTAGTAGTA 1200
 DB 1618 ACTACGACTAAGTTAACAGATGATGCTGCTAGTATTAATTTACGAATCCTGTAGTAGTA 1677
 QY 1201 ACCGGAACAATAGATAATACCGGTAAATGCCAATAAAGTGTGGTAACTTTACCGGAGCA 1260
 DB 1678 ACCGGAACAATAGATAATACCGGTAAATGCCAATAAAGTGTGGTAACTTTACCGGAGCA 1737
 QY 1261 AGTACGCTAACCGGATAATATAGTAAACACGCGAGTATTAGCAGAGGTAAAGGTAGGAGCA 1320
 DB 1738 AGTACGCTAACCGGATAATATAGTAAACACGCGAGTATTAGCAGAGGTAAAGGTAGGAGCA 1797
 QY 1321 GGTGCTGCAAAATACAAAGCGGAGTAGTAAAGCGAATGCAATTAACCTTTAACCGGATAAT 1380
 DB 1798 GGTGCTGCAAAATACAAAGCGGAGTAGTAAAGCGAATGCAATTAACCTTTAACCGGATAAT 1857
 QY 1381 GCGTCACTAGTAACTTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1858 GCGTCACTAGTAACTTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917
 QY 1441 CTCTTCCCAACAGTGAATATAGGAGCAGGATAAATTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 DB 1918 CTCTTCCCAACAGTGAATATAGGAGCAGGATAAATTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1977

```
QY 1501 GCGAATAATATAGATTTTGGAGCTCGAGTAAATTTAGAGTTTAAACGGTCTCCGCGTAAG 1560
|||||
Db 1978 GCGAATAATATAGATTTTGGAGCTCGAGTAAATTTAGAGTTTAAACGGTCTCCGCGTAAG 2037
|||||
QY 1561 AATTAATACTTAATCGGAACACTATAGCAACGGTATATATGCTACACTTAATTAATGCT 1620
|||||
Db 2038 AATTAATACTTAATCGGAACACTATAGCAACGGTATATATGCTACACTTAATTAATGCT 2097
|||||
QY 1621 GCTGCTACAGTGATGCAAAATGATGTTAGTAGCCAGTGGCAAAATTAACATTCAA 1680
|||||
Db 2098 GCTGCTACAGTGATGCAAAATGATGTTAGTAGCCAGTGGCAAAATTAACATTCAA 2157
|||||
QY 1681 AATAATAAGATTTTGTATAAATGCTAAGAACGCTGATGTTGATATATAGACGCTCAG 1740
|||||
Db 2158 AATAATAAGATTTTGTATAAATGCTAAGAACGCTGATGTTGATATATAGACGCTCAG 2217
|||||
QY 1741 GCGATTAGTTTAAAGGAGCAGCTTCACGCTCTTTCTTAGCTAACGTTAGTCTACAGATG 1800
|||||
Db 2218 GCGATTAGTTTAAAGGAGCAGCTTCACGCTCTTTCTTAGCTAACGTTAGTCTACAGATG 2277
|||||
QY 1801 ATAGATTATCACCTCTTAAATAATCATTTACCGGCTCTTGTCTAACCGTGGTGAGTTAA 1860
|||||
Db 2278 ATAGATTATCACCTCTTAAATAATCATTTACCGGCTCTTGTCTAACCGTGGTGAGTTAA 2337
|||||
QY 1861 TTCCTTTGAGTCCAAACAACTTATGACCGTACAGGTGACGTTGGAGCTAAACAAATAG 1920
|||||
Db 2338 TTCCTTTGAGTCCAAACAACTTATGACCGTACAGGTGACGTTGGAGCTAAACAAATAG 2397
|||||

RESULT 2
AE008674/c
LOCUS
DEFINITION Rickettsia conorii Malish 7, section 106 of 114 of the complete genome.
ACCESSION AE008674 AE006914
VERSION AE008674.1 GI:15620409
KEYWORDS
SOURCE
ORGANISM Rickettsia conorii.
REFERENCE
AUTHORS Rickettsia conorii.
Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
1 (sites)
Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E.,
Raoult,D. and Claverie,J.M.
Selfish DNA in protein-coding genes of Rickettsia
Science 290 (5490), 347-350 (2000)
20485642
PUBMED 11030655
2 (sites)
Ogata,H., Audic,S. and Claverie,J.-M.
Selfish DNA and the origin of genes
Science 291 (5502), 252-253 (2001)
3 (bases 1 to 10078)
Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
Raoult,D.
Mechanisms of evolution in Rickettsia conorii and R. prowazekii
Science 293 (5537), 2093-2098 (2001)
21442074
PUBMED 11557893
4 (bases 1 to 10078)
Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
Raoult,D.
Direct Submission
Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean
Moulin, Marseille Cedex 05 13385, France
A public version of R. conorii genome database is accessible at
http://igs-server.cnrs-mrs.fr/. The database intends to provide
updated data.. Annotation of the genome is an ongoing task whose
goal is to make the genome sequence more useful. Comments to the
authors are appreciated.
```

FEATURES
Source

gene

CDS

Location/Qualifiers

```
1..10078
/organism="Rickettsia conorii"
/strain="Malish 7"
/db_xref="taxon:781"
complement(93..6158)
/gene="rmpA"
/note="RC1273"
complement(93..6158)
/gene="rmpA"
/codon_start=1
/transl_table=11
/product="190-KDa cell surface antigen"
/protein_id="AA03811.1"
/db_xref="GI:15620410"
/translation="MANISPKLFOKAIOGQKKAALFTTSTAAIMLSSSGAIGIAVS
GVATNNAAFSDNVGNWNEITAVGNTTPARGPQNNWMAFTYGGDYITADVAADHIIT
AIVNADTPIGLNIAQNTVVGSIIVTGGNLLPVTITAGSLTGLNGNADAANGHGCPA
DNYTGLGNIALGGANAALIIOQAAPAKITLAGNIGGIIITVKDAAITGNTIGNAL
ATVNVGAGIATLEGAI IKATTTKLTNAASVLTLTNNAVLGTADINTGTVDNVGNL
NGALSOVTGNIGNTNALATISVGAGKATLGGAVIKATTKLTNDASAVTFNPVVTG
AIDNTNANNNGIVTGDSTVIGNIGNTNALATISVGAGKATLGGAIKATTKLTND
ASVTFNPVVTG AIDNTNANNNGIVTGDSTVIGNIGNTNALATISVGAGKATLGG
GAIKATTKLTNDASAVTFNPVVTG AIDNTNANNNGIVTGDSTVIGNIGNTNAL
LATISVGAGKATLGGAIKATTKLTNDASAVTFNPVVTG AIDNTNANNNGIVTFT
GDSVTGNIGNTNALATISVGAGKATLGGAIKATTKLTNDASAVTFNPVVTGAI
DNTNANNNGIVTFTGNSVTGNIGNTNALATVNVGAGIATLEGAVIKATTKLTN
AASVLTNNAVLGTADINTGTVDNVGNLNGALSOVTGNIGNTNALATISVGAGKATL
GGAVIKATTKLTNDASAVTFNPVVTG AIDNTNANNNGIVTGDSTVIGNIGNTN
ALATVNVGAGLRRVQGGVKSNTLTNDASAVTFNPVVTG AIDNTNANNNGIVTF
GDSVTGNIGNTNALATISVGAGKATLGGAIKATTKLTNDASAVTFNPVVTG
IDNTNANNNGIVTFTGDSVTGNIGNTNALATVNVGAGTLOAGSSLDANNIDFGARS
TLENGPLDGGGNAIPYFKGAIANGNAILNVTKLTLLTAHLIIGIVAEINIGAGNL
FADASAGDVLTILNAQDILHFRALDSALVLSNLTGVGNVNNILLAADLVAPGDEGTV
FDDVGNLNGISNVAGAAINQDVGNGKFNLTLLIYNAVITDDVLEGIONVLIINNAD
FTSTAFNAGT IQINDATYITDANNGLNIPAGNIKFAHADQILQNSGNDRTITL
GANIDPNDDEGIVILNSVTAGKLTITAGKTGGAKHLODIPKGEGRDCTAGTTFN
TNILVDITGOLELGTANVLEFKDAVOLTOGNTGIFLDFNAKNGTIVTLNNVVA
GTVKTGCTGNTGLIIVLGASNLNRVNGIAMLKVGAGNVTIAGKNVIGELGCTGNT
LTLPAPFKLTGTSINKTGQALKLNFNGSVGVVTAANSVDITTAGTAFSSVSN
AKGTATLGGTTSFAGTNTGAVTLAKGSIISFAKNVTATSFANSATISFNGSLAFN
SNITGSGTTLTLAGANQVYTTGTGTDITLTNTFDGAAGSGGNILKISGSLDLSGV
SNLAVVYATNFMNNTSPDTKYVISAETAGGLKPTPKENVKITINNDRVDFOTFD
ASTLTFAEDIAAGVIDEDPAGGPLANIPNAANTKSLMELEDAPNGSDARAFNE
GLMTPLQEDATTHLMQDVYKPSDTIAAVNVQVYASNISSNITALNARMKVQAGNK
PVSGDEMDAKFGAWISPPVGNATQKMCNSISGYKSDTGTGTTGFGDGFSDLLVLGL
RYATRADTDIKLNNKTDGKNKNESILYGLYSPYENFVEAIASYSDNKIKRSKR
RVATLTLETGYQTANGKYKSESYTGQLMAGYTYMSENILTLPLAGLRYSTIKDKSY
KETGTYTQNLTVKGKNTYTFDGLLAKVSSNINVNEIVLTPELAMVDYAFKNVSAI
DARLQGMTAPLPTNSFKQSKTSFDDVGVGVTAKHKMEYGINVDNIGSKYFAQOGSVK
VRVNF"
complement(6777..9011)
/gene="ftsK"
/note="RC1274"
complement(6777..9011)
/gene="ftsK"
/codon_start=1
/product="cell division protein ftsK homolog"
/transl_table=11
/protein_id="AA03812.1"
/db_xref="GI:15620411"
/translation="MLYINKILSNKKVQAVILGIIGLGIIVILTSYNIDDPSPFSVT
TEYHNLVIGFVSLDCLVQFFGLAFIIPACFVWRCWRCWGRGSRFMRFMVMLL
ALVSSSTLLSKIKLEFIPANAGGAIGIANSFFRNTQNLRLYLLIFFITLIVLVEI
KFTLSLNFIIKLGNFLIYRQSLHNFVLSRLSLPFTPKNDKINTSSYQKPVSEK
VKPEARSVPANPKFSPKSPKISQSEIAPLPSLLRDPKPHVKGASLSLEKQ
KABELTFLANDFGVKGQIININOGVYQVEFEPAAGTKSRVVLGSLDGLARSLALS
TRAVIPGKVLGIELPKOREFEFLKELIETPEYQDKSTLLPLVLGDLQAGPLVAD
LAKMPLHVLVAGTTGSGKSVGINVMVSLRYRTPPECFRTIMDKMLELSAYDGIPLH
LTPVTEPSRAVVALKWKAKEMENRYRMSNIGVKNIAGNAKILEAVKNERIERSI
QTFVDEPTGKIYETVMKMKELPIYIVIVDEADMLVAGKQIEMLIQRLAARAA
GIHINATQRPVSDVITGVIKANFPSPKTSFKTSKIDSRVTLGQSGEQILGMDMLF
MGSTSKISRVHGPVNEAEITEYLKESGTPEYISAVTEQPEEDDDSGIDIGDGTSD
```

```

EVLYKAVOIVRDERKSSYSYIQRSLRIGYNKAANLVEKMEKEGIVSPNHTGKREIL
LPER"
gene complement(9645..9968)
CDS /gene="RC1275"
complement(9645..9968)
/gene="RC1275"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03813.1"
/db_xref="GI:15620412"
/translation="MGVRLLEFALSVAASHSHPLLKFIQITTIIVQSFEMISGFYMA
MIOSNYTSIKIKFWKRYRLPYIFCILDVTFLOQKFSFLSNVIFHRPLVFLIFAN
LTMLL"
BASE COUNT 3031 a 2134 c 1617 g 3296 t
ORIGIN

Query Match 32.1%; Score 615.4; DB 1; Length 10078;
Best Local Similarity 65.7%; Pred. No. 1.3e-103;
Matches 958; Conservative 0; Mismatches 471; Indels 30; Gaps 3;
QY 1 ATGGCGAATATTTCTCTAAATATTTTCAAAAAGCAATTCAAAAGCTCTTAAACCTGCT 60
DB 6158 ATGGCGAATATTTCTCTAAATATTTTCAAAAAGCAATTCAAAAGCTCTTAAAGCGCT 6099
QY 61 TTATTCACCACTCAACCGCAGCGAATAAGCTTAACCGGTAGTGGAGTCTTGGTGTGCA 120
DB 6098 TTATTCACCACTCAACCGCAGCGAATAAGCTGAGTAGTAGCGGGCAGCTCGGTATTGCT 6039
QY 121 AGAACGCTACTGCTGATGTGTCAGAGCTTGCAGCGGACAAATATAGT-----171
DB 6038 GTTTCAGGTGTTATGCTACTAATAATAATGTCAGCATTTAGTAGTGTGGCAATAAT 5979
QY 172 -----CCTGAGCGGTGCTTTTGTAGCGGTTCTACTTTTACAAATATACCGGT 219
DB 5978 TGGAAATGAGATAACGGCTGACGGGTAGCTAATGGTACTCTGCTCGCGTCTCAAAC 5919
QY 220 GCTTTTACGGTTACTGATGCTGACGTAAGTGTGTCGATAGTAAATATAATTTTGA 279
DB 5918 AATTGGGCATTACTTACGGTGTGATTATATCTATCTAGTATGATGATGATGATGATGAT 5859
QY 280 GCAGGTCTTTTTCAGTAACCTGCTGATATTTTATTTAGTGTGCTGATGATGATGATGATGAT 339
DB 5858 ATTACGGCTAATAATGTTGGCGATACCTACCCATAGGTCTAATAATGCTCAAAATACC 5799
QY 340 GCTAATAAAGCTGCAGTAAATATTTGATGATGTTTAACTTAACTTAACTTAACTTAACTTAACT 399
DB 5798 GTCGTTGGTTCGATTAAACGGGAGGTAACTTGTGCTTACTTACTTACTTACTTACTTACTTACT 5739
QY 400 ACTGCACGCTACGGTGCAGAAATCCTGCTGTTTATTCAGGTGGACAAAGCTGCTGTAAT 459
DB 5738 AGCTTAACTTAAACGGTAAATAATGCTGATGCTGCAAAATCATGTTTGTGCTCCTGCC 5679
QY 460 AATACATATACCTGCTTAGTAAATATTAATCTAGTGTGAGCGAATCCGGTGTGCACTATT 519
DB 5678 GATAATATACAGGTTAGGAATATAGCTTTAGGGGAGCGAATGCTGCACTAAATAT- 5620
QY 520 GCTTTCAGATCCAGATGATTAGGACCAATAACGCTTGCAGGAATATAGATGAGGAGGT 579
DB 5619 -----ACAATCTGACGCCCGGCAAGATAAACACTTGCAGCAATATAATGAGGAGGT 5565
QY 580 ATAATACTGACATACAGATGCTGCAATTAACGGAACAAATAGTAACTAATACTAATCCGGCA 639
DB 5564 ATAATACTGCAAGACAGATGCTGCAATTAACGGAACAAATAGTAACTAATACTAATCCGTTA 5505
QY 640 GCTCAATAGCATTTGGAGCAAGTACGCTTCTTCTGGAGGCGCAGTTTATAAGCCACT 699
DB 5504 GCAACAGTGAATGTAGGACAGGTATAGCCACGTTTAGAGGAGCGGATTTAATGAGTACT 5445
QY 700 ACGACTAAGTAAACAAATGACGCGCGGTATTAACTTAACTTAACTTAACTTAACTTAACTTAACT 759
DB 5444 ACGACTAAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5385

```

```

QY 760 ACAGGTGCTGTTGATTAACACACACAGCGGTGATGATGATGCTCTTAAATTTTAAACGGA 819
DB 5384 ACAGGTGCGATTGATAACACACACAGCGGTGATGATGATGCTCTTAAATTTTAAATGCT 5325
QY 820 CGGTTGAGTCAAGTAAGTAACTGGAATATAGTAACTAATAATTAAGTAACTGTA 879
DB 5324 GCATTAAAGTCAAGTAACCGGGAATATAGTAACTAATAATTAAGTAACTGTA 5265
QY 880 GGAGCAGTACGCTACGTTTGGGGAGCGGTTATTAAAGCCACTACGACTAAGTAAAG 939
DB 5264 GGAGCAGTAAAGCCAGCTTGGGGAGCGGTTATTAAAGCTACTAGCACTAAGTAAACA 5205
QY 940 AATGACGCGTGGTATTAACACTTACAAATGCAAGTATTAACAGGTGCTGTTGATAACACC 999
DB 5204 GATAATGCGTCAGCAGTGACATTTACGAATCTGTAGTGTGACCGGAGCATAGATAAT 5145
QY 1000 ACAGGCGGTGATTAATGATGCTGCTAAATCTAAGTGGAGCATTAAGTAAACCGGG 1059
DB 5144 ACCGGTAATGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5088
QY 1060 AATATAGTAACTAATAATTTAGCCAGCATTAATATAGGAGCAGGTGTGGCTACCTTG 1119
DB 5087 AATATAGTAACTAATAATTTAGCCAGCATTAATATAGGAGCAGGTAAAGCCAGCTTA 5028
QY 1120 GATGGAGCGGTTTAAAGCTACTAGCACTAAGTAACTAAGTAACTAAGTAACTAAGTAACTA 1179
DB 5027 GGGGAGCGGATTATTAAGCTACTAGCACTAAGTAACTAAGTAACTAAGTAACTAAGTAACTA 4968
QY 1180 TTTTACGAATCTCTAGTGTAGTAAACCGGAGCATTAATTAAGTAACTAAGTAACTAAGTAA 1239
DB 4967 TTTTACGAATCTCTAGTGTAGTAAACCGGAGCATTAATTAAGTAACTAAGTAACTAAGTAA 4908
QY 1240 GTGCTAATCTTTTACCGGAGCAGTACGCTAAGTAACTAAGTAACTAAGTAACTAAGTAACTA 1299
DB 4907 ATAGTAACTTTTACCGGAGCAGTACGCTAAGTAACTAAGTAACTAAGTAACTAAGTAACTA 4848
QY 1300 GCAGAGTAACTGAGGAGCAGGTTTCTGCTCAATTAACAGCGGAGTAAAGCGAAT 1359
DB 4847 GCCAGCAATAAGTGTAGGAGCAGGTTTACCGGAGCAGTAACTAAGTAACTAAGTAACTAAGTAA 4788
QY 1360 GCAATAAATTTACGGAATATGCTAGTAACTAAGTAACTAAGTAACTAAGTAACTAAGTAACTA 1419
DB 4787 ACAGTAACTTACGGAATATGCTAGTAACTAAGTAACTAAGTAACTAAGTAACTAAGTAACTA 4728
QY 1420 GGTAGTATAGTGTGATACAG 1438
DB 4727 GGAGCGATAGATAATACCG 4709

```

```

RESULT 3
RIRANT190K
LOCUS R.rickettsii 190 kd antigen gene, complete cds.
DEFINITION M31227.1 GI:152465
ACCESSION M31227.1
VERSION 1
KEYWORDS cell surface antigen.
SOURCE R.rickettsii (strain R) DNA, clones p1100 and pGAM21.
ORGANISM Rickettsia rickettsii
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
Rickettsiaceae; Rickettsiinae; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 7088)
AUTHORS Anderson,B., McDonald,G., Jones,D. and Hagnery,R.
TITLE Protective protein antigen of Rickettsia rickettsii has tandemly
repeated: near-identical sequences
JOURNAL Unpublished (1990)
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by B.Anderson, 12-JAN-1990.
FEATURES
source
1..7088
/organism="Rickettsia rickettsii"
/db_xref="taxon:783"
-35_signal 3..8
-10_signal 26..31

```



```

RBS      58. 62
          /note="ribosome binding site"
CDS       70. 6819
          /note="cell surface antigen"
          /codon_start=1
          /transl_table=11
          /protein_id="AAA26380.1"
          /db_xref="GI:152466"
          /translation="MANISPKLFKALIOOGLKAALFTTSTAAMLSGSGALGVATGVI
          AINNNAESNVGNWNNNEITTAAGVANGTPAGGPONNNAFTYGGDYTTADADRRIK
          AINVAGTTPVLNITQNTVVGSIITKGNLIPVTNAGKSLTLNGNNAVAANHGDFAPI
          ATVNGAGTALGGANAALIIOQAAPSKITLAGNIDGGIITVKTDAAINGTIGNINAL
          ATVNGAGTALGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNGVNLN
          NGALSOVTDIGNTSLATISVGAGTATLGGAVIKATTTKLTDRASAVKFTNPVVTC
          AIDNTGNANGIVFTGNTSVGNVNTNALTAVNGAGLLOVGGVYKANTINLTDN
          ASAVTTPNPVVTGALDNTGNANGIVFTGNTSVGNVNTNALTAVNGAGLLOVGGVYK
          GAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNGVNLNNGALSOVTDIGN
          TSLATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVVTCALDNTGNANGIV
          TFGNTSVTDIGNTSLATISVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVL
          TGAIDNTTGGDNGVNLNNGALSOVTDIGNTSLATISVGAGTATLGGAVIKATTTK
          ITNASAVKFTNPVVTCALDNTGNANGIVFTGNTSVTDIGNTSLATISVGAGTATL
          ATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNGVNLNNGALSOV
          DIGNTSLATISVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTT
          DNGVNLNNGALSOVTDIGNTSLATISVGAGTATLGGAVIKATTTKLTNAASVLT
          TNANAVLTGAIDNTTGGDNGVNLNNGALSOVTDIGNTSLATISVGAGTATLGGAV
          IKATTTKLTDAASAVKFTNPVVTCALDNTGNANGIVFTGNTSVTDIGNTSLAT
          VNGAGLLOVGGVYKANTINLTGNASAVTTPNPVVTCALDNTGNANGIVFTGNS
          TVTGNVNTNALTAVNGAGLLOVGGVYKANTINLTGNASAVTTPNPVVTCALDNT
          GNANGIVFTGNTSVTDIGNTSLATISVGAGTATLGGAVIKATTTKLTNAASVLT
          NGPLDGGKALIPYFKGAIANGNNAILNNTKLTASHLTIGTVAEINIGAGNLTID
          ASVGTITLNAQNTINFRADSVLNSLTGKVVNNLILAAADLVAPGADGTVWNGV
          NGLNVSAGTARNIGDGGNKENTLIIYNAVITTDVLEGNLQVNLINKNADPTSS
          TAFNAGAIQINDATYIDANNGLNIPAGNIOFADHQAQVLVQSSNGDRITILGANI
          DPNDDEGIVILNSVTAGKCLITAGGKTFGGAHKQLTILFKGAGCSTAGTIFNTNI
          VLDITQDELECATANVLFNDVOLTQTNIGGLFDNFNANGMVTLNANNVAVAGVQ
          NTGNTNLTILVLAGSLNRYNGTAMLVKVGAGNVI IAKGKVKIGIEIGTNTLITLP
          AHNFTLNTGKQALKNFMNGSVGVGTATNSVGDITTAGATSFASVNAKGT
          ATLGSTTSFANTFTNGAVTLAKGSIITSAKNVTATSEVANSATINFSNLAFTNSIT
          GGGTTLTGANOVTYTGTSFTDLTNTFDGAAGSGNLIKSGSLDLSGVSTLA
          LVVATNFDMNISPDKTYTVISAFTAGLKPTSKENVKIINDNREVDVTFDASTL
          TLFAEDIAADIDGDFAPGKPLANIIPNAANIKKSLMEDAPNGDQKQANFNGLMT
          PLOEDATHLIDVQKPSDTIAVNNQVNASNITNALNMRDKVQSGNKPVS
          GDEMDAFKAWISPFQVKNSTGYKSDPTGGTIGDFGVSDDLALGLAYTR
          ADTDLKLNKNTGKNKYSNIYSLYGLNYPENLFEVETASYSDNRIRSKSRVIA
          TLTETVQTAGNKYKSESYTGOLMAGTYMMPENILNLTPLAGLRYSYTIKDKYKETG
          TTYONLTVKGNKTFDGLLGAKVSSNINVEILVTPELYAMVDYAFNKYSADARL
          OGMTAPLPNSFKQSKTSFDVGVGVGTAKHKMEXRINTDNTNIGSKYFAQGGSVKRVN
          P"
          repeat_region 703..3564
          /note="432 bp repeats"
          misc_signal 6839..6876
          /note="transcription termination signal"
BASE COUNT 2380 a 1125 c 1618 g 1965 t
ORIGIN

```

```

Query Match      31.1%; Score 597.6; DB 1; Length 7088;
Best Local Similarity 63.9%; Pred. No. 2.7e-100;
Matches 982; Conservative 0; Mismatches 518; Indels 36; Gaps 4;

QY 1 ATGGCGCAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGGCTTTAAACTGCT 60
    |||||
DB 70 ATGGCGCAATATTTCTTCAAAATATTTTAAAGCAATTTCAAAAGGCTTTAAAGCGCT 129
    |||||

QY 61 TTATTCACCACCTCAACCCAGCGATAATGCTTAACCGGTAGTGGAGTCCTTGGTCTGCA 120
    |||||
DB 130 TTATTCACCACCTCAACCCAGCGATAATGCTAGTAGTAGTGGGCGACCTCGGTGTGCT 189
    |||||

QY 121 AGAACCGTAACTGCTGATGGTCAGAGCTTGCAGCGGGAACAAATATAGTCTCTGAGGCC 180
    |||||
DB 190 ACAGGTGTATTGCTACTATAATAATGCAGCATTTAGTAATAATCTTGGCAATAATAT 249
    |||||

QY 181 GTGTCTTTTGTAGCGGTTCTACTTACAAATATACCGGTGCTTTTACGGTTTACTGATGCT 240
    |||||

```

```

DB 250 TGGATGAGATAACAGGCTCAGGGGTAGCTAATGTTACTCTCTGGCGGTCTCTCAAAAC 309
QY 241 GACGTAAGTGTTCGTGCATTAGATTTAAATAATTTTTCGACAGGCTCTTTTTCAGTAAT 300
    |||||
DB 310 AATTTGGGCAATTTACTTTACGGTGTGATTATATCTGTCTCAGTACAGCCGATCTGATT 369
    |||||
QY 301 GGTGATATTTCTATTAGTGTTCAGTGTGTAGATACGGGAGGAGCTAATAAATCTTGCAGTAA 360
    |||||
DB 370 ATTAAGGCTATAAATGTTCCGGGTACTACTCCGCTAGGTCTAATAATATTACTCAAAATCT 429
    |||||
QY 361 ATTGATG-----ATGTTTTAACCTTTAACTTTAAACAGGTACCGGT 399
    |||||
DB 430 GTCGTTGGTTCGATTATAACGAAAGTAATCTTCTCCCTGTTACTCTTAATGCCGCAAA 489
    |||||
QY 400 ACTGCAGCCTCAGGTGCAAAATCCTGCGTGTATTTCCTCAAGGTGGGCAACGCTGCTGCTAAT 459
    |||||
DB 490 AGCTTAACTTTAAATGGTATAATGCTGTGCTGCAAAATCATGGTTTGTATGCGCCTGCC 549
    |||||
QY 460 AATACATATCTGCTTTAGGTATAATACTTAGGTGGAGGCAATCCCGTGTGTGACTATT 519
    |||||
DB 550 GATAATTATACAGTTTAGGAATAATAGCTTTAGGGGGAGCGAATGCTGCATAATAT- 608
    |||||
QY 520 GCTTCAGATCCAGATGTATTAGGACCAATAACGCTTTGCAGGAAATATAGATGGAGGAGT 579
    |||||
DB 609 -----ACAATCTGCAGCTCCGTCARAGATAACACTTTCAGGAAATATAGATGGAGGAGT 663
    |||||
QY 580 ATAATAACTGACAAATACAGATGCTGCCATTACGGGAACAATAGGTAAATCTTAATCCGGCA 639
    |||||
DB 664 ATAATAACTGTCAAGACAGATGCTGCCATTACGGGAACAATAGGTAAATCAAAATGCATT 723
    |||||
QY 640 GCTCAAAATAGCATTTGAGCAAGTACGCTTTCTCTTGGAGGGGCGACTTATTAAGGCCACT 699
    |||||
DB 724 GCAACAGTGAATGTAGGACAGCAGTACAGCCACGTTAGGGGGAGCGGTTTATTAAGTACT 783
    |||||
QY 700 ACAGCTAAGTTTAAACAAATCAGCGCGGTATTAACTTTTAACTTACAAATGCAATGCAATGTA 759
    |||||
DB 784 ACAGCTTAAATTAACGAATCTGCTCGGTATTAACTTTTAACTTACAAATGCAATGCAATGTA 843
    |||||
QY 760 ACAGGTGCTGTGTGATAACACACCGCGGTGATGATGATGATGATGATGATGATGATGATGAT 819
    |||||
DB 844 ACAGGTGCGGATGTGATAACACACCGCGGTGATGATGATGATGATGATGATGATGATGATGAT 903
    |||||
QY 820 GCGTGTAGTCAAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 879
    |||||
DB 904 GCATTAACTCAAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 963
    |||||
QY 880 GGAGCAGGTACCGGTACGTTAGGGGAGCGGTTTATTAAGGCCACTACGACTAAGTTAAACG 939
    |||||
DB 964 GGAGCAGGTACAGCCACGTTAGGGGAGCGGTTTATTAAGGCCACTACGACTAAGTTTGA 1023
    |||||
QY 940 AATGCAGCGCTCGGTATTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 999
    |||||
DB 1024 GATGCTGCTCAGCAGTGAATTTACGAATCTCTGTAGTGGTACTGGAGCGGATAGATAAT 1083
    |||||
QY 1000 ACAGCGGTGATAAATAGTGTGCTGTAATCTTAAAGTGGAGCATTTAGTCAAGTAACTTAACT 1059
    |||||
DB 1084 ACCGGTAAATGCAAAATAATGTTAGTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1140
    |||||
QY 1060 AATATAGTAAATACAAATTCATTAGCCAGATAAATAATAGGACGAGGTGGTGGTACTCTG 1119
    |||||
DB 1141 AATGTAGTAAATACAAATTCATTAGCCAGATAAATAATAGGACGAGGTGGTGGTACTCTG 1200
    |||||
QY 1120 GATGAGCGGTTTATTAAGCTACTACGACTAAGTTTAACTTAACTTAACTTAACTTAACT 1179
    |||||
DB 1201 CRAAGTGGAGTGTAAACCAATAACAAATTAACCTTAACTTAACTTAACTTAACTTAACT 1260
    |||||
QY 1180 TTTACGAATCTCTGTAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1239
    |||||
DB 1261 TTTACGAATCTCTGTAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1320
    |||||
QY 1240 GTGGTAACTTTTACCGGACAGTACGGTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1299
    |||||
DB 1321 ATAGTAACGTTTACCGGTAATAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1380
    |||||

```


Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.
 1 (bases 1 to 6361)
 Croquet-Valdes, P.A., Weiss, K. and Walker, D.H.
 Sequence analysis of the 190-kDa antigen-encoding gene of
 Rickettsia conorii (Malish 7 strain)
 Gene. 140 (1), 115-119 (1994)
 94171067
 PUBMED
 8125327
 2 (bases 1 to 6361)
 Croquet-Valdes, P.A.
 Direct Submission
 Submitted (23-AUG-1993) P.A. Croquet-Valdes, University of Texas
 Medical Branch at Galveston, Pathology, 11th Street and Texas
 Avenue, Galveston, Texas 77555, USA
 Location/Qualifiers
 1. 6361
 /organism="Rickettsia conorii"
 /strain="Malish 7"
 /db_xref="taxon:781"
 3. 8
 26. 31
 59. 63
 72. 6137
 /codon_start=1
 /transl_table=11
 /product="190-kDa antigen (rOmpA)"
 /protein_id="AA17405.1"
 /db_xref="GI:467814"
 /translation="MANISPKLFKKAIOQGLKAAFTSTTAAILMSSGALGIASVGV
 IATNNAAFSNVGNNEITPAAGVNGTPARQPNNAFTYGGDYTVADVAHHIT
 ENVADPTPIGLNAQNTVGVSIITGNLPLVTINAGKSLITLNGNDAANHGADPA
 DNYTGLGIALGANAALIIOQAPAKITLAGNNGGILIVKTDAAINGTIGNAL
 ATVNGAGIATLEGAIIKATTKLTNAASVLTINVRALFGLADINTGYDNLVNL
 NGALSOYTGNGTGNALATISVGAGKATLGAVIKATTKLTIDNASVTFNPVVV
 ADINTGNNGIVTIGDSTVGTGNIGTGNALATVNGAGLQVGVVYKANTINLND
 ASAVTFNPVVVGAIDNTGNNGIVTIGDSTVGTGNIGTGNALATISVGAGKATLG
 GAIITKATTKLTIDNASVTFNPVVVGAIDNTGNNGIVTIGDSTVGTGNIGTGN
 LATISVGAGKATLGAIKATTKLTIDNASVTFNPVVVGAIDNTGNNGIVTIGDST
 VGTGNIGTGNALATISVGAGKATLGAIKATTKLTIDNASVTFNPVVVGAIDNT
 GDSVTFNPVVVGAIDNTGNNGIVTIGDSTVGTGNIGTGNALATISVGAGKATLG
 VLTINVRALFGLADINTGYDNLVNLGALSOYTGNGTGNALATISVGAGKATLG
 GGAVIKATTKLTIDNASVTFNPVVVGAIDNTGNNGIVTIGDSTVGTGNIGTGN
 ALATISVGAGKATLGAIKATTKLTIDNASVTFNPVVVGAIDNTGNNGIVTIGD
 TGDSTVGTGNIGTGNALATISVGAGKATLGAIKATTKLTIDNASVTFNPVVVGA
 IDNTGNNGIVTIGDSTVGTGNIGTGNALATISVGAGKATLGAIKATTKLTIDNAS
 TLEFNPGLDGGKAIPIYKGAIAANGNALINVTKSLTASHLITIGVFAEINIGAGNL
 FAIDNASGDTVILNAODIHFALDSALVSLNLTGVGNVNLILADLVAPGVDEGTVP
 DGVNGLNIGSNVAGARNIGDGVGNKFNLTLLIYNAVITDDVNLGTONVLLNNAD
 FTSSTFAGNTIQINDATYITDANGNLINPAGNIOFAHADAOLILONSSGNDITFL
 GANIDPDNDDEGIIVILNSVTAGKLTITAGKTGGAGHKLQDIVEFGDGFDTAGTTFN
 TTNITVDITGOLELGTATNVVLFKDAVLTQTNIGTGNIGTGNALATISVGAGKATLG
 GTVNTGTNGTILVLGASNLVRNGIAMLKVGAGNVITAKGNGVKIGELQGTNT
 LTLPAHFKLTSIKNTGGOALKLNFNGGSSGVVGTAAANSVGDITTAGATSPASSVN
 AKGTALGCTTSFANTFNTGAVTLAKGSIISFAKNVTATSFVANSATINFGNSLAFN
 SNITGSGDTLFGANQVYITGTFDTLTLNFTFDGAAGKSGNLLKSGTDLGSLV
 SNLALVYATFNMDNNSPDPTKYVISAETAGLKPPTKENVKITINNDNFVDFTFD
 ASTLTLEAFDAIDVDFAPGGLPANIANAIIKSLMELEDPAGNSDARQAFNE
 PLVMSDEMDAKFAGNISPPVGNATKMCNISISYKSDTTGGTIGTGFVSDDLVLGL
 GVSSGDEMDAKFAGNISPPVGNATKMCNISISYKSDTTGGTIGTGFVSDDLVLGL
 AYTRADIDILKNNKGTGKNKVESNIYSLGYLSPVENLFEVIAISYSDNKKISKSR
 RVIAITLETGVYGFANGKYESYTGQMLVGYPPYMLPENVLPLAGLRYSTIKDKSY
 KETGTYTNTKGNKNTNFDGLGKAVSSNINVAIPLTPELVAMVDYAFKKNVSAT
 DARLQGITPPLSTNSFKQSKTSFDVGVGVTAHKMMVEYRINYDINIGSKYFAQOGSVK
 VRVNF"
 repeat_region 705..2882
 /rpt_type=tandem
 terminator 6156..6194
 BASE COUNT 2143 a 1015 c 1431 g 1772 t
 ORIGIN
 Query Match 30.6%; Score 587.8; DB 1; Length 6361;
 Best Local Similarity 64.9%; Pred. No. 1.7e-98;

Matches 947; Conservative 0; Mismatches 482; Indels 30; Gaps 4;
 QY 1 ATGGCGAATATTTCTCTAAATTTATTTCAAAAGCAATTTCAAAAGGCTTTAAACCTGCT 60
 DB 72 ATGGCGAATATTTCTCTAAATTTATTTCAAAAGCAATTTCAAAAGGCTTTAAACCTGCT 131
 QY 61 TTATTCCACCCTCAACCCGAGGATATGCTTAACCGGTAGTGGAGTCTTCTGGTCTGCA 120
 DB 132 TTATTCCACCCTCAACCCGAGGATATGCTTAACCGGTAGTGGAGTCTTCTGGTCTGCA 191
 QY 121 AGAACCGTAACCTGCTGAGGCTTGCAGCGGAGCAAAATATAGTCTCTGGAGCC 180
 DB 192 GTTTCAGGTGTTTACTACTATTAATATGAGCATTTAGTGATAATTTGGCATATAT 251
 QY 181 GGTGCTTTTGTAGCGGTTTCTACTTTACAAATACCGGTCTTTTACGGTTTACTGA --- 236
 DB 252 TGGAAATGAGATAACGGCTGCAGGGGTAGCTAATGTGTACTCTCTCGGGGCTCTCAAAAC 311
 QY 237 -----TGCTGAGCTAAGTGTCTGTCATAGATTAAATAATTTTTCAGCAGCT 285
 DB 312 AATTGGGCATTTACTTACGGTGTGATTAATGCTGCTGACAGCTGACCGCATATAT 371
 QY 286 CTTTTTTCAGTAACCTGGTGATATTTCAATAGGTTGAGTACGAGGAGGAGTAAAT 345
 DB 372 ATTACGGCTATAAATGTTGCGGATATCTCTCCATAGGTCTTAATATTTCTCAAAATACC 431
 QY 346 AAACCTTGAGTTAATATGATGATGCT-----TTAACCTTAACTTTAACAGCTACCGGT 399
 DB 432 GTCTGTTGCTGATTAACGGGAGGTAACTTGTTCCTGTTTACTATTAAATCCCGGCAA 491
 QY 400 ACTGCAGCTACCGTGCAAACTCCTGCTGTTTATTCCTCAAGGTGGCAAGCTCTGCTAAT 459
 DB 492 ACCTTAACTTTAAACCGTAAATGCTGATGCTGCAAACTATGCTTGTGATGCTCTGCTGCC 551
 QY 460 AATACATATCTGCTTTTAGGTAAATAACTCTAGTGGAGCAATACCGGTTTACTATTT 519
 DB 552 GATAATATACAGTTTAGGAAATATAGCTTTAGGGGAGCAATGCTGCATTAATAT - 610
 QY 520 GCTTCAGATCCAGATGATTAGGACCAATAACGCTTGCAGGAAATATAGATGAGGAGGT 579
 DB 611 -----ACATCTGACGCCCCGCAAGATACACCTTGACGCAATATAAATGGAGAGGT 665
 QY 580 ATAATACTGACATACAGATGCTGCCATTAACGGAACAATAAGTAATACTAATCCGCA 639
 DB 666 ATAATACTGTCACAGCAGATGCTGCCATTAACGGAACAATAAGTAATACTAATCCGTTA 725
 QY 640 GCTCAAAATAGCATTTGGAGCAAGTACGCTTCTCTTGGAGGGGAGTATTAATAGCCACT 699
 DB 726 GCAACAGTGAATGTAGGAGCAGGTATAGCCAGCTTAGAGGGGCGGATTTAATAGCTACT 785
 QY 700 ACGACTAAGTTAAACAAATGTCAGCGCGGTATTAAACACTTACAAATGCAAAATCAGTATTA 759
 DB 786 ACGACTAATTAACGAATGCTGCTGCGGTATTAAACCTTACAAATGTAATCGAGTATTA 845
 QY 760 ACAGTGTCTGTTTAACACACACGCGGTGATGATGATGCTTAAATTTAAATCCGCA 819
 DB 846 ACAGTGTCTGTTTAACACACACGCGGTGATGATGATGCTTAAATTTAAATCCGCT 905
 QY 820 GCTTTCAGTCAAGTAACTGGAAATATAGTATACAAATTAATGACGCAATATAGTATGA 879
 DB 906 GCATTAAGTCAAGTAACTGGAAATATAGTATACAAATTAATGACGCAATATAGTATGA 965
 QY 880 GGAGCAGGTACGGCTACGTTAGGGGAGCGGTATTAAAGCCACTACGACTAAGTAAAG 939
 DB 966 GGAGCAGGTAAAGCCAGCTTAGGGGAGCGGTATTAAAGCTTACTACGACTAAGTAAAG 1025
 QY 940 AATGAGCGGTGTTTAAACACTTACAAATGAGTATTAACAGGCTGCTGTTGATACACC 999
 DB 1026 GATAATGCTGAGCAGTGCACATTTACGATCTCTGAGTGGTACCGGAGGATATAT 1085
 QY 1000 ACAGCGGTGATATGATGAGTGTCTGTAATCTAATGAGTGGAGCATGAGTCAAGTACCGG 1059
 DB 1086 ACCGTAATGCAAAATATGATAGTAACTGTTACCGGTG---ATAGTACAGTAAGTGGC 1142

QY	1060	AATATAGGTAATACAAATTCATTAGCCACGATAAATATAGGACGAGTGTGGCTACCTTG	1119
Db	1143	AATATAGGTAATACAAATTCATTAGCAACAGTGAATGAGGACGAGCTTTGCTACAAAGTA	1202
QY	1120	GATGGACGGTTATTAAAGCTACTACGACCTAAGTTTACAGCATGATCGCTCAGTATTTAATA	1179
Db	1203	CAAGGTGGAGTGGTAAAGCAAAATACAAATAACCTTACCGGATATGCGTCAGCAGTGACA	1262
QY	11180	TTTACGAATCCTGTAGTAGTAACACGGGAGCAATAGATAATACCCGGTAATGCCAATAAAAGGT	1239
Db	1263	TTTACGAATCCTGTAGTGGTACCGGAGCGCATAGATAATACCCGGTAATGCCAATAATAGGT	1322
QY	1240	GTGGTAATCTTTACCGGAGCAAGTACGGTACCGGATATATAGGTAAACACGGCAGTATTA	1299
Db	1323	ATAGTACGTTTTACCGGTGTAGTACAGTAACTGGCAATATAGTAAATACAAATGGCATTA	1382
QY	1300	GCAGAGGTAACCGTAGCAGCAGAGTTTCCCTGCAAAATACAAAGCGGAGTAGTAAAGCGCAAT	1359
Db	1383	GCCACGATAAGCTAGCAGGACAGGTAAAGCCACGGTTAGGGGGAGCGATTATTTAAAGCTACT	1442
QY	1360	GCAATAAACCTTAAACGGATAATCGCTCAGTAGTACATTTTACCCGGTGATAGTACCGTAACA	1419
Db	1443	ACGACTAAGTTAACGGATATCGCTCAGCAGTGACATTTACGAAATCCTGTAGTGGTGACC	1502
QY	1420	GGTAGTTAGGTGGTACAG	1438
Db	1503	GGAGCGATAGATATACCG	1521

RESULT 6	
AF210694	
LOCUS	575 bp DNA linear BCT 11-JAN-2001
DEFINITION	Rickettsia sp. California 2 outer membrane protein OmpA (ompA)
	gene, partial cds.
ACCESSION	AF210694
VERSION	AF210694.1 GI:12082488
KEYWORDS	
SOURCE	Rickettsia sp. California 2.
ORGANISM	Rickettsia sp. California 2.
	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
	Rickettsiaceae; Rickettsiae; Rickettsia.
REFERENCE	1 (bases 1 to 575)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

```

FEATURES             Location/Qualifiers
     1..575
     /organism="Rickettsia sp. California 2"
     /strain="California 2"
     /db_xref="taxon:147259"
     gene
     1..575
     /gene="OmpA"
     CDS
     1..575
     /gene="OmpA"
     /codon_start=1
     /transl_table=1
     /product="outer membrane protein OmpA"
     /protein_id="AAG48555.1"
     /db_xref="GI:12082489"
     /translation="LFQKAIQKGLKTALETTAAIMLTGSGVLGAARTVTADGAELA
AGTNGRGAGAFVAGTGYLTGATFTVDADVSRALDNNFAAGLSEFTGDISLGSVV
DTGCKANLAVNIIDGDLTLITGTGTAAYGANPALLFQGGQAAANNYYTALGNITLIGA
NAGLTITADPDVLGPTITLAGNIDGGITDNT"
     164..104 C 134 a 173 t
BASE COUNT
164 a

```

Query Match 29.98; Score 575; DB 1; Length 575;

[illegible]

RESULT 7	
RIROMPAB	
LOCUS	3087 bp DNA linear BCT 13-MAY-1996
DEFINITION	Rickettsia conorii outer membrane protein A (ompA) gene, partial cds and repeat region.
ACCESSION	L01462 L01463
VERSION	L01462.1 GI:152485
KEYWORDS	outer membrane protein A; repeat region.
SOURCE	Rickettsia conorii (strain Kenya tick typhus) DNA.
ORGANISM	Rickettsia conorii
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsia; Rickettsia; spotted fever group.
AUTHORS	Gilmore, R.D. Jr.
TITLE	Comparison of the romPA gene repeat regions of Rickettsiae reveals species-specific arrangements of individual repeating units
JOURNAL	Gene 125 (1), 97-102 (1993)
MEDLINE	93194085
PUBMED	7680636
FEATURES	Location/Qualifiers
source	1..3087
gene	/organism="Rickettsia conorii"
	/strain="Kenya tick typhus"
	/db_xref="taxon:781"
	1..3087
	/gene="ompA"
	<1..>3087
	/gene="ompA"
	/codon_start=1
CDS	

```
/transl_table=11
/product="outer membrane protein A"
/protein_id="AA09908.1"
/db_xref="GI:152486"
/translation="IGNNALRTVNVGAGIATLEGALIKATTKTLTNAASVLILTNV
AVLGAIDNTGVDNVLNGLSQTGNIGNTALATISVGAGKATLGGAVIKAT
TKTLTNDASQVFTNPVVTGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISV
AGKATLGGALIKATTKTLTNDASQVFTNPVVTGAIDNTGNANNGIIVFTGDSVTG
NIGNTALATISVGAGKATLGGALIKATTKTLTNDASQVFTNPVVTGAIDNTGNAN
NGIIVFTGDSVTGNIGNTALATISVGAGKATLGGALIKATTKTLTNDASQVFTNP
VVVGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISVGAGKATLGGALIKATTK
KLTDNASQVFTNPVVTGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISVNGAG
LLRVGGVFSNTINLTNDASQVFTNPVVTGAIDNTGNANNGIIVFTGDSVTGNIG
NNTALATISVGAGKATLGGALIKATTKTLTNDASQVFTNPVVTGAIDNTGNAN
IIVFTGDSVTGNIGNTALATISVGAGKATLGGALIKATTKTLTNDASQVFTNPV
VTGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISVGAGKATLGGALIKATTK
TNDASQVFTNPVVTGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISVNGAGTA
TLEGAVIKATTKTLTNAASVLILTNVAVLGAIDNTGVDNVLNGLSQTGNIGN
IGNTALATISVGAGKATLGGAVIKATTKTLTNDASQVFTNPVVTGAIDNTGNAN
GIATFGDSVTGNIGNTALATISVGAGKATLGGALIKATTKTLTNDASQVFTNPV
VVTGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISVGAGKATLGGALIKATTK
LTDNASQVFTNPVVTGAIDNTGNANNGIIVFTGDSVTGNIGNTALATISVNGAG"
repeat_region 1..>3087
/note="regions 443..2170 and 2396..3043 are composed of
repeat units 216 base pairs long"
/rpt_type=tandem
/rpt_unit=443..658
1..225
repeat_region 226..442
repeat_region 443..658
/partial
/gene="ompA"
repeat_region 2171..2395
repeat_region 3044..>3087
BASE COUNT 1077 a 474 c 785 g 751 t
```

```
ORIGIN
Query Match 24.6%; Score 473.2; DB 1; Length 3087;
Best Local Similarity 74.9%; Pred. No. 2.5e-77;
Matches 650; Conservative 0; Mismatches 203; Indels 15; Gaps 4;
QY 577 GGTATTAATACATGACATGCTGCCATTAACGGACAATAGTAACTAATCGG 636
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1912-GGTATAGTAACTTTACCGGTATAGTACAGTAATAGGTAATACAAATGCA 1971
QY 637 GCAGCTCAATAGCATTCGACAGTACGCTTCTCTTGGAGGGGCGAGTTAAGCC 696
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1972 TTAGCAACAGTGAATGTAGGACGAGTTTGTCTACGAGTACAAGCGGAGTGTAATCG 2031
QY 697 ACTGACCTAACTTAAACAAATCAGCGCGGTATTAAACACTTACAAATGCAATGCGATGA 756
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2032 AATACAAATAACTTAACGGATAATCGTCACAGGTGACATTTACGATCC---TGTAGTG 2088
QY 757 TTAACAGGTGCTGTGATTAACACACAGCGGTGATGATGTAGGTCTCTTAATTTAAAC 816
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2089 GTACCGGAGCGATAGATAATACC---GGTAATGCAAAATAGGTATAGTAACTGTTACC 2145
QY 817 CGAGCGTTGAGTCAAGTAACTGAATATAGCTAATACAAATTCATAGCGACAATAAGT 876
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2146 GG---TAATAGTACAGTAACTGGCAATATAGGTAATACAAATGCAATACGATGAAT 2202
QY 877 GTAGGACGAGTACGGCTACGTTAGGGGAGCGGTATTATTAAGCCACTACGACTAAGTTA 936
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2203 GTAGGACGAGTATAGCCAGCTTAGAGGAGCGGTATTATTAAGCTACTACGACTAATTA 2262
QY 937 ACCAATGACGCGTGTATTAACACTTAC-----AAATGCAGTATTAAACAGGTGCTGTT 990
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2263 ACCAATGCTGCTGCGTATTAAACCTTACAAATGTAATAGTATTAACAGGTGCGATT 2322
QY 991 GATAACACACAGCGGTGATTAATGTAGTGTCTGTAATCTAAGTGGAGCATTTAGTCAA 1050
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2323 GATAACACACAGCGGTGATTAATGTAGTGTCTTAAATTTAAATTTAAGTGTGATTAAGTCAA 2382
```

```
QY 1051 GTAAACCGGGAATATAGGTAATATACAAATTCATTAGCCAGTAAATATAGGACGAGGTGTG 1110
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2383 GTAAACCGGGAATATAGGTAATATACAAATTCATTAGCCAGTAAATATAGGACGAGTAA 2442
QY 1111 GCTACCTTGGATGGAGCGGTTATTAAAGCTACTAGGACTAAGTTAACAGATGATCGGTCA 1170
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2443 GCCAGCTTAGGGGAGCGGTTATTAAAGCTACTAGGACTAAGTTAACAGATGATCGGTCA 2502
QY 1171 GTATTAATATTACGAATCCCTGTAGTAGTAACCGGAGCAATAGATAATACCGGTAAATGCC 1230
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2503 CAGGTGACATTTACGAATCCCTGTAGTAGTACCGGAGCGATAGATAATACCGGTAAATGCC 2562
QY 1231 AATAAGGTGTGGTAATCTTTACCGGAGCAATAGTACCGTAATATATAGGTAAACAG 1290
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2563 AATAATGTTATAGGCACCTTTACCGGTGATAGTACAGTAACGTGCAATATAGGTAAATACA 2622
QY 1291 SCAGTATTAGCAGAGGTAAAGCTAGGAGCAGGTTTGTCTGCAATACACAGCGGAGTAGTA 1350
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2623 AATGCATTTAGCAACAGTGAATGTAGGAGCAGGTTTGTCTGCAATACACAGCGGAGTAGTA 2682
QY 1351 AAAGCAATGCAATAAATCTTAACGGATTAATCGTCAGTAGTAACATTTACCGGTGATAGT 1410.
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2683 AATGCATATACATAAATCTTAACGGATTAATCGTCAGTAGTACATTTACGAATCCTGTA 2742
QY 1411 ACGGTAAACAGGTAGTATAGGTGGTACAG 1438
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2743 GTGGTGACCGGAGCGATAGATAATACCG 2770
RESULT 8
AF231136 3184 bp DNA linear BCT 02-JAN-2001
LOCUS Rickettsia sp. California-2 nonfunctional protein ompA gene,
DEFINITION partial sequence.
ACCESSION AF231136
VERSION AF231136.1 GI:12005024
KEYWORDS Rickettsia sp. California-2.
ORGANISM Rickettsia sp. California-2.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 3184)
AUTHORS Raoult,D.
TITLE Description of a new Rickettsia, Rickettsia pulicis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3184)
AUTHORS Roux,V. and Raoult,D.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Unite des Rickettsies, Faculte de Medecine,
27 Boulevard Jean Moulin, Marseille 13385, France
FEATURES
Location/Qualifiers
source 1..3184
/organism="Rickettsia sp. California-2"
/strain="California 2"
/db_xref="taxon:135271"
<1..>3184
/misc_feature /note="nonfunctional protein ompA due to mutation during
evolution"
BASE COUNT 1054 a 547 c 673 g 910 t
ORIGIN
Query Match 23.5%; Score 452; DB 1; Length 3184;
Best Local Similarity 100.0%; Pred. No. 1.9e-73;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1469 GAATAACATTACGACCGGAGGAGCCCTAGCTGCGAATAATATAGATTTTGAGCTGCGA 1528
Db 1 GAATAACATTACGACCGGAGGAGCCCTAGCTGCGAATAATATAGATTTTGAGCTGCGA 60
QY 1529 GTAATTTAGATTTTAACGGTCTCGGGTAAGAAATTAACTTAATCGGAACATATAGCAA 1588
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GTAATTTAGATTTTAACGGTCTCGGGTAAGAAATTAACTTAATCGGAACATATAGCAA 120
```

Qy 1589 ACGGTAATTAAGTACACCTTAATATTAATGCTGCTGCTACAGTGAATGCAATGATGTTA 1648
 Db 121 ACGGTAATTAAGTACACCTTAATATTAATGCTGCTGCTACAGTGAATGCAATGATGTTA 180
 Qy 1649 GTATAGGCACAGTGCACAAATTAACATTCAAATTAAGATTTTGTATAAATGCTA 1708
 Db 181 GTATAGGCACAGTGCACAAATTAACATTCAAATTAAGATTTTGTATAAATGCTA 240
 Qy 1709 AGAAGCTGATGTTGATATATTAGACGCTCAGCGGATAGTTTAAAGGACGCTTCAC 1768
 Db 241 AGAAGCTGATGTTGATATATTAGACGCTCAGCGGATAGTTTAAAGGACGCTTCAC 300
 Qy 1769 GTCCTTTCTTACGTAACGTTAGTCTACAGATGATAGATTTATCACTCTTAAATATCATTT 1828
 Db 301 GTCCTTTCTTACGTAACGTTAGTCTACAGATGATAGATTTATCACTCTTAAATATCATTT 360
 Qy 1829 ACCCGGCTTCGTAACGCTGCTGCTGAGTTAATTCCTTTTGGTACCAACAACTTATGAC 1888
 Db 361 ACCCGGCTTCGTAACGCTGCTGCTGAGTTAATTCCTTTTGGTACCAACAACTTATGAC 420
 Qy 1889 CGTAAAGGTGACGTTGGAGCTAAACAAATAG 1920
 Db 421 CGTAAAGGTGACGTTGGAGCTAAACAAATAG 452

RESULT 9
 AFI49108
 LOCUS
 DEFINITION Rickettsia australis strain PHS outer membrane protein A (ompA)
 ACCESION AFI49108
 VERSION AFI49108.1 GI:5081770
 KEYWORDS Rickettsia australis.
 SOURCE Rickettsia australis.
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
 1 (bases 1 to 7067)
 Stenos, J. and Walker, D. H.
 The rickettsial outer-membrane protein A and B genes of Rickettsia
 australis, the most divergent rickettsia of the spotted fever group
 Int. J. Syst. Evol. Microbiol. 50 Pt 5, 1775-1779 (2000)
 20487299
 11034486
 2 (bases 1 to 7067)
 Stenos, J. and Walker, D.
 Direct Submission
 Submitted (10-MAY-1999), Australian Rickettsial Reference
 Laboratory, Douglas Hocking Medical Institute, Geelong Hospital,
 Ryrie St., Geelong, VIC 3220, Australia
 Location/Qualifiers
 1. 7067
 /organism="Rickettsia australis"
 /strain="PHS"
 /db_xref="taxon:787"
 354. 6939
 /gene="ompA"
 354. 359
 /gene="ompA"
 377. 382
 /gene="ompA"
 409. 413
 /gene="ompA"
 421. 6741
 /gene="ompA"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein A"
 /protein_id="AAD39531.2"
 /db_xref="GI:11641393"
 /translation="WANTSPKLFQKAIHSLKAALEFTTAAIMLSGGAGVAAAGVI
 SYNDAAFSNRAVANNWEITAGGAANGHADGPQDNEATPYGNNHTITADEAGRIITA
 INVAGTTPVALNSTQNTSVGSIVTGGNLLPVTIADGKSLITLGTAKVAADHGFDAAD

NYTGLGAIAGGATAGLTIOSATPALINLAGAIDGDNNDHGEITVTRTSFTGVIQGT
 KTLHATFNNRAGGYSTAGASATTTVMIGEDAGNVASVVOGLINGAVNFAADGALT
 ANNGIAGAVTTANNNTGTLTVGAGDVTGAIVNGGNILKOVLFNGSNVATIDATNVT
 INNAANVTAAAGATAAVNFAADGALTANNAGVATTTANNNTGTLTVGAGDVTGAIV
 VNGNLTGKOVLFNGSNVATIDATNVTINNAANVTAAAGATAAVNFAADGALTANNAG
 IAGVATTTANNNTGTLTVGAGDVTGAIVNGGNILKOVLFNGSNVATIDATNVTINNV
 AANVTAAAGATAAVNFAADGALTANNAGVATTTANNNTGTLTVGAGDVTGAIVNGG
 NILKOVLFNGSNVATIDATNVTINNAANVTAAAGATAAVNFAADGALTANNAGVAT
 VTTANNNTGTLTVGAGDVTGAIVNGGNILKOVLFNGSNVATIDATNVTINNVAAV
 TAAGATAAVNFAADGALTANNAGVATTTANNNTGTLTVGAGDVTGAIVNGGNILK
 OVLFNGSNVATIDATNVTINNAANVTAAAGATAAVNFAADGALTANNAGIAGVATTA
 NNTGTLTVGAGDVTGAIVNGGNILKOVLFNGSNVATIDATNVTINNAANVTAAAG
 AITAAVFAADGALTANNAGVATTTANNNTGTLTVGAGDVTGAIVNGSNILKOVLF
 NGASNVTIDATNVTINNVAAVTAAGATAAVNFAADGALTANNAGITAVTAAGONG
 TLTVAGVGTGAGTGAASGLKVLNAKVDNGAADLVTSIYARTVNFDDIAGGAVGT
 IQVGGIATNVNFGGNAOGSLTLNGPVGKSYTLSTGTANGNANTINNTLGLVTFAB
 NAXIGTVAQINTODNKTFAINVKYADIEILNAQADIFGANSKLFVLNNSATDDRVVT
 INKDLPAFATGGMLLFTANLITLQGDGAKAIGTAGNKLASLNLGKGFAPNNITDT
 TNYLAFNIINTTQFVDVGGTITQISVINIQAANASIGLPAAGASYTIDANGVNGILA
 NGOTINFAHEDAEVLQNSAAGNSTITLNAALDPGAPNGKGLAVDSXAGGRVILASV
 RNATYCTAVNKLKLEFRNGTFOITDIFANNLELPVPAATYKNDIKXNLSFGAAT
 ALTQNGINGVDFNQAAVILGANKNITGVSVTSSNGVNGITITAGSTINGHITNI
 AMUKGAGAVSITGGGTSITEIQNGTALLTLNPNFTGINKTGGOALLKLNFTNG
 GSVRGVYTAANSVGDITTTGTTFEASSYNAKGAATLCGTTSFADTFTTGAATLAKA
 SITNEAKNVTATSEFANNATINFGNSLAFNSNITGSGTTLTGASQVTTGTSFDT
 LTLNTPFDGAASDGHILIKSGSTLDLSGVSTLALVVTATNFDINISPTKVTLSA
 ETAGLKPMPAGNVKVTVNDNRVNFVTFDESTLTLFAEDIAEEVEEDPKGGLAN
 INNAANVKSLEMEQAPITGSDACLAFDNWKLNPIQADAVTHLMQDVVKPSDITAA
 INQVLSNISSSLNLNARMADIAQAAVAAAGDEDRDAFGAWISFPVGNAMQKMRNN
 ISGYKSDTGGTIGFDGLVSDDLALGAYTRADTDIKLKNKKNKYESHYSLYG
 LYNVPYENLCFKAIASYSNKRIRSKRRRIATALDIVGNOTAGSKYKSESYTGQLMAG
 YTYMPENISLPLAGRLYSATDKGYKETGTTNONTLVKGNYSYTFDGLGAKVSSN
 INVGEVLTPEFVAVDYAFKNKVPADARLOGMTSPFTNFSKQSKTSFDICVGITA
 KHMMEYGINYDTNIGSKYFAQGGVKVRVNE"
 1216..3318
 repeat_region /rpt_type=direct
 /rpt_unit=1216..1470
 6903..6939
 /gene="ompA"
 BASE COUNT 2327 a 1216 c 1615 g 1900 t 9 others
 ORIGIN
 Query Match 13.5%; Score 259; DB 1; Length 7067;
 Best Local Similarity 79.4%; Pred. No. 4.9e-38;
 Matches 312; Conservative 5; Mismatches 75; Indels 1; Gaps 1;
 Qy 1529 GTAATTAGAGTTTAAAGGTCCTGCGGTGAAGATTTATTAATCGGAACTATAGCAA 1588
 Db 3572 GTACTTTAATTTGAACGGTCCTGCGGTGAAGATTTATTAATCGGAACTATAGCAA 3631
 Qy 1589 ACGGTAATTAAGTACACCTTAATATTAATGCTGCTGCTACAGTGAATGCAATGATGTTA 1648
 Db 3632 ACGGTAATTAAGTACACCTTAATATTAATGCTGCTGCTGCTACAGTGAATGCAATGCTA 3691
 Qy 1649 GTATAGGCACAGTGCACAAATTAACATTCAAATTAAGATTTTGTATAAATGCTA 1708
 Db 3692 KTATCGGTACGGTGTGACAAATTAACATTCAGATATAAGACTTTTGTCAATAATGTTA 3751
 Qy 1709 AGAAGCTGATGTTGATATATTAGACGCTCAGCGGATAGTTTAAAGGACGCTTCAC 1768
 Db 3752 AGAAGCTGATGTTGATATATTAGACGCTCAGCGGATAGTTTAAAGGACGCTTCAC 3811
 Qy 1769 GTCCTTTCTTACGTAACGTTAGTCTACAGATGATAGATTTATCACTCTTAAATATCATTT 1827
 Db 3812 AGCTTTTTTTAGTTAAACAATAGTGTACAGATGATAGATTTATCACTCTTAAATATCATTT 3871
 Qy 1828 TACCCGGCTCTTGTACGCTGGTGGTGGAGTTAATTCCTTTTGGTACCAACAACTTATGA 1887
 Db 3872 TACCCGGCTCTTGTACGCTGGTGGTGGAGTTAATTCCTTTTGGTACCAACAACTTATGA 3931
 Qy 1888 CCGTCAAGGTGACGTTGGAGCTAAACAAATAG 1920
 Db 3932 CTTTCAAGGTGATGCTGCASCTAAGGCAATAG 3964

RESULT 10
 RMU83439 3167 bp DNA linear BCT 10-SEP-1998
 LOCUS
 Definition
 Rickettsia mongolotimonae rompa (ompA) gene, partial cds.
 accession
 U83439.1 GI:1778884
 version
 1
 keywords
 Rickettsia mongolotimonae.
 ORGANISM
 Rickettsia mongolotimonae
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
 REFERENCE
 1 (bases 1 to 3167)
 Fournier, P.E., Roux, V. and Raoult, D.
 Phylogenetic analysis of spotted fever group rickettsiae by study
 of the outer surface protein rompA
 JOURNAL
 Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
 MEDLINE
 98404549
 PUBMED
 9734038
 REFERENCE
 2 (bases 1 to 3167)
 Raoult, D., Fournier, P.-E. and Roux, V.
 Direct Submission
 JOURNAL
 Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
 CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
 FEATURES
 Location/Qualifiers
 1..3167
 /organism="Rickettsia mongolotimonae"
 /strain="HA-91"
 /db_xref="taxon:45261"
 1..3167
 /gene="ompA"
 <1..>3167
 /gene="ompA"
 /codon_start=3
 /transl_table=11
 /product="rOmpA"
 /protein_id="AAC35175.1"
 /db_xref="GI:1778885"
 /translation="ITLHAGSGLDANNIDFGARSTLEFNGGGIPYVFKAIAANGNNA
 ILNVTGKLLTAHYLTIGTVAEINAGNIFAIDASGDVTLNAQIHFRALDSTLL
 SLNLTGVGNLILALDPADEGTVFNGVNGNLINSGVAGTARNIGDGGKFNFL
 ALLIYNAVYITDDVLEQGVNLINNNADFTSSTAFNAGAIQIONATYITIDANNGLN
 IPAGNTQFAHADAQLILKSSGNDRTITLCANIDPDNDDEGIVILNSVTAGKRLTAG
 KTGFGAGKPTDFTVFGAGDCSTAGTFTNIVLDITGOLEGATATNVLPFKDVA
 LKTCGTGIGGLDFENAKGTVLNNVNVAGTVQNTGTTNGTTLVLGASNLNRVNGIA
 MLKVGAGNTYIAKGNVKGIEQCTGCTNLTLPAPHLNLTGSKNTKGQALKNFMNGG
 SVSGVGTAAANSVGDITTAGTASFASVNAKGTAKLGGTTSFAHTTFTGAVTLAKGS
 ITSFANKVITSEFVANSINFGNSAFNSNITGSGTTLTLAGANQVYITGTGVTFTLL
 TLNTPDGAAGSGNLLIKSGSTLDLSGYSALVATNFATNMNNISPTPKYTVIAE
 TVGGLKPTKENVKITINDNREVDFTDASTLTFLAKDIAADVIDKDFAPGPGPLTA
 PNAQNTKSLKLMEDAPNSDAPAFNFGMLTPOEADATTHLMODVVKPSPITAAV
 NNOVYASNISSNTALNARMKDVQAKNGPVSQSGDEMDAKFGAWLSPFVGNATQMK
 NSLSGTGDTGTGTFDGVSDDLVLGLATRADTDILKLNKNTGDKNKVESNIYSL
 YGLYSVPENLFEALASYSDNKIRSKSRVVIATTLTGVGTANGKYKSESYTGOLM
 AGYVMPNENINILPLVGLRYSTIKDKYKETGTTQNLTVKGNNTYTFDGLLGAKVS
 SINNVNEIVLTPELYAMVDYAFKNVSAIDARLQGTATPLPTNSFKQSKTSPFDVGVG
 TAKYKMEYGINVYDNTGSKYFAQQ"
 BASE COUNT 1062 a 513 c 666 g 926 t
 ORIGIN
 Query Match 8.1%; Score 154.6; DB 1; Length 3167;
 Best Local Similarity 69.8%; Pred. No. 9.4e-19;
 Matches 224; Conservative 0; Mismatches 94; Indels 3; Gaps 1;
 QY 1469 GATAACATTCAGAGCGGAGGAAGCCCTAGCTGCGAATATATAGATTTTGGAGCTCGCA 1528
 Db 1 GTATAACATTACAGCTGGAGGAAGTCTAGATGCGAATAATATAGATTTTGGAGCAGGA 60
 QY 1529 GTAATTTAGAGTTTAAAGGTCCTCCCGTAGAGATTATAACTTAATCGGAACATAGCAA 1588
 Db 61 GTACTTTAGAGTTTAAAGGAGGTTGGTGTATCCCTTATTTATTTTAAAGGAGCTATAGCAA 120

QY	181	GGTCTTTTGTAGCGGGTCTTACTTTTCAATATACCGGTGCTTTTACGGTTTACTGATGCT	240
Db	181	TGGGATGAGATAACCGCTGAAGGGGTAGCTAATGGTATTCTCTGCTGGCGGTCTCAAAAC	240
QY	241	GACGTAAGTGTTCGTGCTAGATTTAAATTAATTTTGGACGAGTCT-----	287
Db	241	AATTGGGCATTCTTACTTACAGTGTGATATATACTATCATCTGCAGATGTAGTCGATCGTATT	300
QY	288	-----TTTTTTCAGTAACTGGTGATATTTCATTAGTTTTCAGTGGTAGATACGGGAGGA	339
Db	301	ATTACGGCTAATATGTTGGGGTACTACTCCCGTAGGTCTAATATATGCTCAAAATACC	360
QY	340	GCTAATAAAGTTGCGATTATATATTGATGATGGTTTAAACCTTTAAACAGGTACCGGT	399
Db	361	GTCGTTGGTTCGATATAAAGTATAGAGTAACCTTGTTCGCTGTACTATTGCTGGCAAAAGC	420
QY	400	ACTCAGCCTACGGTGCAGAAATCCGCGTGTATTATCCAAAGTGGACAGCTGCTGCTAAT	459
Db	421	TTAACCTTAAAGGTACTATGCTGTGTGCTCAATCATGGTTTGTATGCTCCGGCCGAT	480
QY	460	AATACATATCTGTTTAGGTAATATATAACTCTAGGTGGAGCGAATGCGGTTTGACTATT	519
Db	481	AAT---TATACAGGTTTAGGAATATAAATCTTTAGGGGAGGAGTCTGAATTAATATA	537
QY	520	GCTTCAGATCCAGATGTATTAGGACCAATACGCTTTCAGGAGGAGTATAGATGGAGGAGT	579
Db	538	CAATCTGCAACCCCGGCAAG-----ATAACACTTCGAGAAATATAGATGGAGGAGT	591
QY	580	ATAATACTGACATACAGATGCTGCCATTAAACGAAC	617
Db	592	ATAATACTGCAAGACAGATGCTGCCATTAAACGAAC	629
RESULT 12			
AF179366			
LOCUS	AF179366	3168 bp	DNA linear BCT 12-SEP-2000
DEFINITION	Rickettsia sp. HL-93 strain HL-93	OmpA (ompA) gene, partial cds.	
ACCESSION	AF179366		
VERSION	AF179366.1	GI:9789177	
KEYWORDS			
SOURCE	Rickettsia hulinii.		
ORGANISM	Rickettsia hulinii		
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;		
AUTHORS	Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.		
TITLE	1 (bases 1 to 3168)		
JOURNAL	Zhang, J.Z., Fan, M.Y., Wu, Y.M., Fournier, P.E., Roux, V. and Raoult, D.		
MEDLINE	Genetic classification of 'Rickettsia hellongjiangii' and		
PUBMED	'Rickettsia hulinii', two Chinese spotted fever group rickettsiae		
REFERENCE	J. Clin. Microbiol. 38 (9), 3498-3501 (2000)		
AUTHORS	20440687		
TITLE	10970415		
JOURNAL	2 (bases 1 to 3168)		
MEDLINE	Zhang, J.Z. and Raoult, D.		
PUBMED	Direct Submission		
REFERENCE	Submitted (18-AUG-1999) Unite des Rickettsies, Faculte de Medecine,		
AUTHORS	27 Bd. Jean Moulin, Marseille 13385, France		
TITLE	Location/Qualifiers		
JOURNAL	1. .3168		
MEDLINE	/organism="Rickettsia hulinii"		
PUBMED	/strain="HL-93"		
REFERENCE	/db_xref="taxon:121193"		
AUTHORS	<1. >3168		
TITLE	/gene="ompA"		
JOURNAL	<1. >3168		
MEDLINE	/genes="ompA"		
PUBMED	/note="rOmpA"		
REFERENCE	/codon_start=3		
AUTHORS	/transl_table=1		
TITLE	/product="OmpA"		
JOURNAL	/protein_id="AAF98268.1"		
MEDLINE	/db_xref="GI:9789178"		
PUBMED	/translation="LHYTUGSLDANNIDFGARSTLEFNGGGGIPYFKGAIANGNNA		

RESULT 14
RSU83455

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 08:37:09 ; Search time 899 Seconds
(without alignments)
4809.607 Million cell updates/sec

Title: US-09-800-065-1

Perfect score: 1920

Sequence: 1 atggcggaattctctaaa.....cgttgagctaaacaataag 1920

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1920	100.0	5513	22	AA18232
2	63.8	3.3	486	22	AAF75507
3	59.2	3.1	510	22	ABA48451
4	59.2	3.1	510	22	ABA66355
5	59.2	3.1	510	22	ABA33417
6	59.2	3.1	510	22	ABA14772
7	59.2	3.1	510	22	AAK40512
8	59.2	3.1	510	22	AAI46546
9	57.6	3.0	4197	16	AAQ99430
					Rickettsia felis o
					Polyglutamine trac
					Human foetal cell
					Human foetal liver
					Probe #11883 for g
					Human brain expres
					Human bone marrow
					Probe #15232 used
					B. sphaericus SLP

c 10	54.2	2.8	14704	13	AAQ20685
c 11	54	2.8	360	23	ABL25081
c 12	54	2.8	2682	23	ABL25080
c 13	53.2	2.8	8201	21	AAA88864
c 14	53.2	2.8	8201	21	AAA88864
c 15	53.2	2.8	8201	24	ABQ73537
c 16	53.2	2.8	8201	24	ABQ73537
c 17	51.6	2.7	1388	21	AAC59039
c 18	51.6	2.7	1388	24	ABQ54638
c 19	51.6	2.7	2556	22	AAI59455
c 20	51.4	2.7	8155	18	AAV74374
c 21	51	2.7	10266	17	AAAT33007
c 22	50.8	2.6	3263	23	ABL10403
c 23	50.8	2.6	16962	23	ABL10402
c 24	50.4	2.6	1923	23	ABL28551
c 25	50.4	2.6	2215	24	ABI99688
c 26	50.4	2.6	2614	24	ABK36100
c 27	50.4	2.6	6604	24	ABN95250
c 28	50.4	2.6	6788	24	ABK83506
c 29	50.4	2.6	6794	21	AAZ45597
c 30	50	2.6	2240	23	ABL28548
c 31	50	2.6	4324	23	ABL28550
c 32	49.6	2.6	2744	16	AAQ98470
c 33	49.4	2.6	7104	23	AAS31998
c 34	49.4	2.6	7107	23	AAS44654
c 35	49.4	2.6	7434	23	AAS52179
c 36	49.4	2.6	7437	23	AAS55232
c 37	48.8	2.5	1037	21	AA59242
c 38	48.8	2.5	1159	21	AA59240
c 39	48.8	2.5	1472	21	AA59241
c 40	48.4	2.5	5120	22	AAC84677
c 41	48.4	2.5	5397	23	ABL29757
c 42	48.4	2.5	7029	23	ABL29756
c 43	47.8	2.5	4590	22	AAH24065
c 44	47.8	2.5	6033	21	AA70152
c 45	47.4	2.5	457	22	ABA44433

ALIGNMENTS

RESULT 1
AA18232
ID AA18232 standard; DNA; 5513 BP.
XX
AC AA18232;
XX
DT 18-DEC-2001 (first entry)
XX
DE Rickettsia felis outer membrane protein genomic DNA (rompA).
XX
KW Outer membrane protein; rompA; antibacterial; ds.
XX
OS Rickettsia felis.

XX Key Location/Qualifiers
FT CDS 478..2334
FT /*tag= a
FT /product= "Rickettsia felis outer membrane protein"
FT /note= "CDS is specifically claimed in claim 4 of
the specification"

PN WO200166691-A2.

XX 13-SEP-2001.

XX 06-MAR-2001; 2001WO-US07820.

XX 06-MAR-2000; 2000US-187323P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Bouyer DH, Crocquet-Valdes P, Stenos J, Walker DH;

applicant

XX WPI; 2001-596829/67.
 DR P-PSDB; AAE10890.
 XX
 XX Novel isolated Rickettsia felis outer membrane protein and
 PT polynucleotide for modulating expression of the protein in a host cell
 XX
 XX Example 1; Page 64-68; 73pp; English.
 XX
 XX The present invention relates to an isolated Rickettsia felis outer
 CC membrane protein (rompA) and its polynucleotide. The rompA gene is
 CC useful for modulating expression of the protein in a host cell. The
 CC rompA gene is useful as probe or for the design of primers to obtain
 CC DNA encoding the protein by either cloning and colony/plaque
 CC hybridisation or amplification using PCR. The present sequence is
 CC Rickettsia felis outer membrane protein genomic DNA (rompA).
 XX
 XX Sequence 5513 BP; 1814 A; 961 C; 1181 G; 1557 T; 0 other;
 SQ
 Query Match 100.0%; Score 1920; DB 22; Length 5513;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCGGAATATTCCTAAATATTTCAAAAAGCAATTCAAAAGGCTTTAAAACCTGCT 60
 DB 478 ATGCGGAATATTCCTAAATATTTCAAAAAGCAATTCAAAAGGCTTTAAAACCTGCT 537
 QY 61 TTATTACCACCTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCCCTGGTGCA 120
 DB 538 TTATTACCACCTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCCCTGGTGCA 597
 QY 121 AGAACCGTAAGTGTGATGTGTCAGAGCTTGCAGCCGGAACAATATAGTCTCTGGAGCC 180
 DB 598 AGAACCGTAAGTGTGATGTGTCAGAGCTTGCAGCCGGAACAATATAGTCTCTGGAGCC 657
 QY 181 GGTCTTTTGTAGCGGCTTCTACTTTTCAATATATACCGGTGCTTTACGGTTACTGATGCT 240
 DB 658 GGTCTTTTGTAGCGGCTTCTACTTTTCAATATATACCGGTGCTTTACGGTTACTGATGCT 717
 QY 241 GACGTAAGTGTCTGCTAGTATTAATTAATTTTTCAGCAGGCTCTTTTTCAGTAACCT 300
 DB 718 GACGTAAGTGTCTGCTAGTATTAATTAATTTTTCAGCAGGCTCTTTTTCAGTAACCT 777
 QY 301 GGTGATATTTCAATAGTGTTCAGTGTAGATACGGGAGGAGCTAATAAAGTTCAGTTAAT 360
 DB 778 GGTGATATTTCAATAGTGTTCAGTGTAGATACGGGAGGAGCTAATAAAGTTCAGTTAAT 837
 QY 361 ATTGATGATGTTTAACTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 420
 DB 838 ATTGATGATGTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 897
 QY 421 CCTGCGTTGTTATCCAAAGTGGCAAGCTGCTGCTAATAATACATATACATCTGCTTTAGGT 480
 DB 898 CCTGCGTTGTTATCCAAAGTGGCAAGCTGCTGCTAATAATACATATACATCTGCTTTAGGT 957
 QY 481 AATATACTCTAGTGTGAGCAATGCGGTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTT 540
 DB 958 AATATACTCTAGTGTGAGCAATGCGGTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTT 1017
 QY 541 GGACCAATAACGCTTGCAGGAATAATAGATGGAGGAGGTATTAATTAACCTGACATACAGAT 600
 DB 1018 GGACCAATAACGCTTGCAGGAATAATAGATGGAGGAGGTATTAATTAACCTGACATACAGAT 1077
 QY 601 GCTGCCAATTAACGAACAATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
 DB 1078 GCTGCCAATTAACGAACAATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1137
 QY 661 AGTACGCTTCTCTTGGAGGGGAGGATTTTAAAGCCCACTACGACTAGTTTAAACAATGCA 720
 DB 1138 AGTACGCTTCTCTTGGAGGGGAGGATTTTAAAGCCCACTACGACTAGTTTAAACAATGCA 1197
 QY 721 CGCCCGGTATTAACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 780

DB 1198 GCGCCGGTATTAAACACTTAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1257
 QY 781 ACAGCGCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 1258 ACAGCGCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
 QY 841 AATATAGGTATATACAAATTCATTAGCGACATATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 1318 AATATAGGTATATACAAATTCATTAGCGACATATGATGATGATGATGATGATGATGATGATGATGAT 1377
 QY 901 GGGGAGCGGTTATTAAAGCCACTTACGACTTAAGTTAAAGTTAAAGTTAAAGTTAAAGTTAAAGTT 960
 DB 1378 GGGGAGCGGTTATTAAAGCCACTTACGACTTAAGTTAAAGTTAAAGTTAAAGTTAAAGTTAAAGTT 1437
 QY 961 CTTTAAATGCGAGTATTAAAGCGGCTGTTGATTAACACACAGCGCGGTGATATGATGATGATGAT 1020
 DB 1438 CTTTAAATGCGAGTATTAAAGCGGCTGTTGATTAACACACAGCGCGGTGATATGATGATGATGAT 1497
 QY 1021 GTCGTAATCTAAGTGGAGCATTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1080
 DB 1498 GTCGTAATCTAAGTGGAGCATTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1557
 QY 1081 TTAGCCACGATAAATATAGGAGCAGGTGTGGCTACCTTGGATGGAGCGGTTATTAAAGCT 1140
 DB 1558 TTAGCCACGATAAATATAGGAGCAGGTGTGGCTACCTTGGATGGAGCGGTTATTAAAGCT 1617
 QY 1141 ACTACGACTAAGTTAAACAGATGATGCTGCTAGTATTAATTTACGAACTCTGTAGTAGTA 1200
 DB 1618 ACTACGACTAAGTTAAACAGATGATGCTGCTAGTATTAATTTACGAACTCTGTAGTAGTA 1677
 QY 1201 ACCGGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGTGGTAAATCTTTACCGGAGCA 1260
 DB 1678 ACCGGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGTGGTAAATCTTTACCGGAGCA 1737
 QY 1261 AGTACGTTACCGATAATATAGGTAAACACCGGAGTATTTAGCAGAGGTAAAGGTAGGAGCA 1320
 DB 1738 AGTACGTTACCGATAATATAGGTAAACACCGGAGTATTTAGCAGAGGTAAAGGTAGGAGCA 1797
 QY 1321 GGTTCGCTGCAATACAAAGCGGAGTAGTAAAGCGAATCAATTAACCTTTAAAGGATAAT 1380
 DB 1798 GGTTCGCTGCAATACAAAGCGGAGTAGTAAAGCGAATCAATTAACCTTTAAAGGATAAT 1857
 QY 1381 GCGTACGTAGTACATTTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1858 GCGTACGTAGTACATTTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917
 QY 1441 CTCTTCGCAACAGTGAATATAGGAGCAGGAGTAACTTACGAGCGGAGGAGCGCTAGCT 1500
 DB 1918 CTCTTCGCAACAGTGAATATAGGAGCAGGAGTAACTTACGAGCGGAGGAGCGCTAGCT 1977
 QY 1501 GCGAATAATATAGATTTTGGAGCTGGAGTAAATTTAGAGTTTAAAGCTCTCGCGGTAAG 1560
 DB 1978 GCGAATAATATAGATTTTGGAGCTGGAGTAAATTTAGAGTTTAAAGCTCTCGCGGTAAG 2037
 QY 1561 AATTAACCTTAATCGGAATATAGCAACCGGTAAATCTACCTAATATATTAATGCT 1620
 DB 2038 AATTAACCTTAATCGGAATATAGCAACCGGTAAATCTACCTAATATATTAATGCT 2097
 QY 1621 GCTGGTACAGTGTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 DB 2098 GCTGGTACAGTGTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2157
 QY 1681 AATTAAGATTTTGTAAATGCTTAAGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 DB 2158 AATTAAGATTTTGTAAATGCTTAAGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2217
 QY 1741 GCGATTAAGTTTAAAGGAGCAGCTTACGCTCTTTTCTTAGCTAACGCTTAGCTACAGATG 1800
 DB 2218 GCGATTAAGTTTAAAGGAGCAGCTTACGCTCTTTTCTTAGCTAACGCTTAGCTACAGATG 2277
 QY 1801 ATAGAGTTATCACTCTTAAATCAATTTACCGGCTCTTGTCAACGGTGGTGGTGGTGGTGGTGGTGGT 1860

Db 2278 ATAGACTTATCACTCTTAAATAATCATTTACCGGCTTCTGCTAACGGTGGTGGTAA 2337
 QY 1861 TTCTTTTGGTCCAAACAACTTATGACCGTACAGGTGACGTGAGCTAAACAATAG 1920
 Db 2338 TTCTTTTGGTCCAAACAACTTATGACCGTACAGGTGACGTGAGCTAAACAATAG 2397

RESULT 2
 AAF75507/c
 ID AAF75507 standard; DNA; 486 BP.
 XX
 AC AAF75507;
 XX
 DT 09-MAY-2001 (first entry)
 XX
 DE Polyglutamine tract coding sequence #2.
 XX
 KW Animal model; polyglutamine tract; neurodegenerative disorder; HDJ1;
 KW heat shock protein 40; tetrairicopeptide repeat protein 2; TPR2; stroke;
 KW myeloid leukaemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
 KW Parkinson's disease; CJD; BSE; Huntington's disease; head trauma;
 KW cancer; ds.
 XX
 OS Drosophila sp.
 XX
 PN WO200112238-A1.
 XX
 PD 22-FEB-2001.
 XX
 PF 14-AUG-2000; 2000WO-US22496.
 XX
 PR 12-AUG-1999; 9805-0148933.
 PR 12-AUG-1999; 9905-0148934.
 PR 18-JAN-2000; 2000US-0177047.
 PR 19-MAY-2000; 2000US-0205720.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Benzer S, Kazemi-Esfarjani P;
 XX
 DR WPI; 2001-147537/15.
 XX
 PT Identifying genes or other compounds that modulate polyglutamine
 PT toxicity, useful for treating Alzheimer's disease, Parkinson's disease
 PT and Creutzfeldt-Jakob disease -
 XX
 PS Disclosure; Fig 1A; 275pp; English.
 XX
 CC The present invention describes a method of screening for genes which
 CC modulate polyglutamine toxicity using animal models with polyglutamine
 CC sequences that cause toxicity in the animal. The model is preferably
 CC Drosophila, and toxic polyglutamine sequences include the human and
 CC Drosophila heat shock protein 40/HDJ1, tetrairicopeptide repeat protein 2
 CC (TPR2) and myeloid leukaemia factor 1 (MLF) genes. The model is useful
 CC for identifying treatments for neurodegenerative and proliferative
 CC disorders, including Alzheimer's disease, Parkinson's disease,
 CC Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
 CC (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar
 CC ataxias, dentatorubropallidolysian atrophy, Kennedy's disease, stroke,
 CC head trauma and cancer.
 XX
 SQ Sequence 486 BP; 193 A; 164 C; 108 G; 21 T; 0 other;

Query Match 3.3%; Score 63.8; DB 22; Length 486;
 Best Local Similarity 48.2%; Pred. No. 4e-06;
 Matches 179; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 110 TTGGTCTGCAAGAACCGTAACTGTGATGTCAGAGCTTGCAGCGCGCAACAATATAG 169
 Db 441 TTGCTGCTCTGTGTGTGCTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 382
 QY 170 GTCCGTGACCGGCTGCTTTTGTAGCGGGTTCACCTTACAATATACCGTGTTCACG 229

Db 381 TTGCTGTTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 332
 QY 230 TTACTGATGCTGACGTAACTGTTGCTGCAATAGATTAATTTTGCAGAGGCTCTTT 289
 Db 321 CTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 262
 QY 290 TTTCAGTAACTGCTGATATTTTCATTAGGTTCACTGCTAGATACGGGAGGAGCTAATAAAC 349
 Db 261 CTGCTGTTGCTGTTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGTTG 202
 QY 350 TTGCAGTTAATATTGATGATGGTTTAACTTAACTTTAAGTACCGGTACTGACGACCT 409
 Db 201 TTGCTGCTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 142
 QY 410 ACGGTGCAAACTCCTGCTGTTGTTATTCCAAAGTGGACAAGCTGCTGCTAATAATACATA 469
 Db 141 TTGTTGCTGCTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 82
 QY 470 CTGCTTTAGGT 480
 Db 81 CTGCTGTTGCT 71

RESULT 3
 ABA48451/c
 ID ABA48451 standard; DNA; 510 BP.
 XX
 AC ABA48451;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #7146.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 4; SEQ ID NO 7146; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene

Db	210	CAATGCAGATGCTGATGCTGGTGTAGTGTGATGCTGATGCTGGTGTAGTGTGATGCTGGTGCAGATG	151	
Qy	464	CATATACTGCTTTTAGGTAATATAACTCTAGGTGAGAGCAATGCCGGTTTGACTATTGCTT	523	
Db	150	TTGATGCAGATGTTGAAGCGGCTTGTGATGCTGATGCAGGTGCTGAAGCAGGTATTGGTG	91	
Qy	524	CAGATCCAGATG	535	
Db	90	CTGATACAGATG	79	
RESULT 6				
AAK14772/c				
ID	AAK14772 standard; DNA; 510 BP.			
XX	AAK14772;			
XX	05-NOV-2001 (first entry)			
XX	Human brain expressed single exon probe SEQ ID NO: 14763.			
XX	Human: brain expressed exon; gene expression analysis; probe;			
KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;			
KW	epilepsy; cancer; ss.			
XX	Homo sapiens.			
XX	OS			
XX	W0200157275-A2.			
XX	09-AUG-2001.			
XX	30-JAN-2001; 2001WO-US00667.			
XX	04-FEB-2000; 2000US-0180312.			
PR	26-MAY-2000; 2000US-0207456.			
PR	30-JUN-2000; 2000US-0608408.			
PR	03-AUG-2000; 2000US-0632366.			
PR	21-SEP-2000; 2000US-0234687.			
PR	27-SEP-2000; 2000US-0236359.			
PR	04-OCT-2000; 2000GB-0024263.			
XX	(MOLE-) MOLECULAR DYNAMICS INC.			
XX	Penn SG, Hanzel DK, Chen W, Rank DR;			
XX	WPI; 2001-483446/52.			
XX	Single exon nucleic acid probes for analyzing gene expression in human brains			
PT	Example 4; SEQ ID NO: 14763; 650pp + Sequence Listing; English.			
PS	The present invention provides a number of single exon nucleic acid			
XX	probes which are derived from genomic sequences expressed in the human			
CC	brain. They can be used to measure gene expression in brain cell samples,			
CC	which may enable the diagnosis and improved treatment of nervous system			
CC	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,			
CC	epilepsy and cancers. The present sequence is one of the probes of the			
CC	invention.			
XX	Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;			
Qy	Query Match 3.1%; Score 59.2; DB 22; Length 510;			
Db	Best Local Similarity 46.1%; Pred. No. 5.8e-05;			
XX	Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;			
Qy	104 GAGTCCTTGGTGTGCTGAAGAACCCTAACTGCTGATGCTGACAGATTTCAGCCGGAACAA			163
Db	510 GTGTGTGCTGTATGCAGGTACCGATGCTGAAGCAGGTGTTGATCCAGGTGCTGAAAAG			451
Qy	164 ATATAGGTTCCTGGAGCCGGTGTCTTTTGTAGCGGTTCTTACATATATACCGGTGCTT			223
Db	450 GTGTGTGGTGGAGGTGCAGATGCTGATGCTGTGATGCTGAATGCTGATGCAGGCGCTGATGCTG			391


```

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta
XX
XX Claim 25; SEQ ID NO 15232; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;
SQ
Query Match 3.1%; Score 59.2; DB 22; Length 510;
Best Local Similarity 46.1%; Pred. No. 5.8e-05;
Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;
QY 104 GAGTCTTGGTCTGCAAGACCGTAACCTGCTGATGGTGCAGAGCTTGCAGCCGGAACAA 163
DB 510 GTGTTGATGCTGATCGCAGGTACCGATGCTGAAGCAGGTGTTGATCGAGTGTGAAAG 451
QY 164 ATATAGTCTCGAGCCGGTGTCTTTGTAGCGGGTCTTACTTTACAATATACCGTGTCTT 223
DB 450 GTGTTGGTGCAGGTGCAGATGCTGATGCTGTGTAGTAATGCTGATCGAGCGCTGATGCTG 391
QY 224 TTACGGTTACTGATGCTGACGTAAGTGTGCTGCAATAGATTAAATAATTTTGCAGCAG 283
DB 390 ATGCAGGTATGTTGTTGTCAGATGTTGATGTCAGATGTTGAAGCAGGTGCTGAAAGAG 331
QY 284 GTCCTTTTTCAGTACTGCTGATATTTTCAATAGTTTCACTGATGCTGATGCGGAGGCTA 343
DB 330 GTGTGATCAGGTGCAGATGCTGATGCTGTAGTGTGATGCTGATGCTGATGCTGTG 271
QY 344 ATAACTTCAGTAAATATGATGATGTTTAACTTAACTTAACTTAACTTAACTTAACTTAA 403
DB 270 CAGATGTTGATGATGATGTTGAAGCAGGTGTTGATGTCAGGTGCTGAAAGTGTGTTGATG 211
QY 404 CAGCCTACGTCGAATCTCTGTTGTTTATTCACAGGTGCAAGCTGCTGCTAATAATA 463
DB 210 CAAGTGCAGATGCTGATGCTGTGTAGTGTGATGCTGATGCTGATGCTGATGCTGATG 151
QY 464 CATATCTCTTTAGTAAATATACTCTAGGTGAGCGAATCCGGTCTTTGACTATTGCTT 523
DB 150 TTGATCAGATGTTGAAGCGGGTGTGTGATGCTGATGCTGATGCTGATGCTGATGCTG 91
QY 524 CAGATCCAGATG 535
DB 90 CTGATACAGATG 79
RESULT 9
AAQ99430
ID AAQ99430 standard; DNA; 4197 BP.
XX
XX AAQ99430;
XX
XX 22-DEC-1995 (first entry)
XX
XX B. sphaericus SLP gene.
XX
XX Surface layer protein; SLP; fusion protein; vaccine; antigen;
KW surface expression; epitope; ds.
XX
XX Bacillus sphaericus.
XX
XX Key Location/Qualifiers
FH 79..85
FT RBS
FT a
FT *tag=
FT CDS 95..3853
FT b
FT *tag=
FT sig_peptide 95..184
FT c
FT *tag=

```

```

mat_peptide 185..3850
/*tag= d
XX
XX WO9519371-A2.
XX
XX 20-JUL-1995.
XX
XX 13-JAN-1995; 95WO-EP00147.
XX
XX 14-JAN-1994; 94GB-0000650.
XX
XX (SOLV ) SOLVAY SA.
XX
XX Deblaere RY, Desomer J, Dhaese P;
XX
XX WPI: 1995-263827/34.
XX
XX P-PSDB; AAR80530.
XX
XX Host cell expressing surface layer protein fusion protein - used for
XX host presentation of antigens and vaccine prodn.
XX
XX Disclosure; Fig. 6; 95pp; English.
XX
XX A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
XX p-13855) surface layer protein was used to screen an HindIII-
XX generated library to isolate the slp gene. Promoter regions
XX of the gene are used in genetic constructs providing surface
XX expression of heterologous proteins in P-1 hosts.
XX
XX Sequence 4197 BP; 1470 A; 731 C; 763 G; 1233 T; 0 other;
SQ
Query Match 3.0%; Score 57.6; DB 16; Length 4197;
Best Local Similarity 47.0%; Pred. No. 0.00032;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;
QY 116 CTGCAAGAACCGTAACTGCTGATGTTGTCAGAGCTTGCAGCCGGAACAAATATAGTCTCTG 175
DB 2919 CTTTAGGAATTTCAATAGCTGATGATGATGCTTAATGTAAGTGCACAACTGTTGATAGTG 2978
QY 176 GAGCCGGTCTTTTGTAGCGGGTCTTACTTTACAATATACCGTGTCTTTTACGGTTTACG 235
DB 2979 CAACGTGTTTCATAAAG-----ATAGTGCATTAATTAATCTTATCTTTACATTAAGT 3032
QY 236 ATGCTGACCTCAAGTGTTCGTCATTAAGTAAATTAATTTTGCAGCAGGTCTTTTTCAG 295
DB 3033 AAACGTGCTATACAGGTGATTTGCTTACAACTTCAAGCTGCTACATATATCTTCTT 3092
QY 296 TAACTGGTGATATTTTCATTAAGTTCAGTGTGATGATACGGGAGGAGCTAATAAATTCGAG 355
DB 3093 TAACTGCTGTATACATTAACAGTTTACTTATGTCAGATGCTTAAAGATGCTGAGGTGTTGCTG 3152
QY 356 TTAATATTTGATGATGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT 415
DB 3153 AAAATATTACTGCTAGCTGAACATTAAGAAAAC---TACTGGAGCAATTTACTTCTGATA 3209
QY 416 CAAATCTGCTGTTGTTTATTTCCAGGTGGACAAAGCTGCTGCTAATAATACATATACCTGCTT 475
DB 3210 CATTTACACAAGGTGATTAACCATCAGCAGCTACAGCAGCTGATATCTTCTTAATCAA 3269
QY 476 TAGGTAATACTCTTAGTGGAGCGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG 535
DB 3270 TTGCTGCAGATTATACATTTTGCACAGGTGAAGGATTCACCTTTAAATATTTGATAATGCTG 3329
QY 536 TATTAGGACCAATAAGCTTGCAGGAAATATAGATGAGGAGGATATAATAAATACATGACAATA 595
DB 3330 GTGCTCAAGTAATTAATCTAGCAGGTAAAGGTTAAAGGTTCAAGGT---GTAGCTGATGCTA 3386
QY 596 CAGATGCTGCATTAACGGAACAAATAGGTAAATTAATTAATTAATTAATTAATTAATTAAT 655
DB 3387 TCAATGCTACATTTTGCAGGTGCTGCAACTGTTTCTGGAGACAAGTAGTATTATTAATCAG. 3446
QY 656 GAGCAAGTAGCTTTTCTTTGGAGGGGAGCTTTATTAAGCCACTAGCACTAAGTTAAACA 715

```


CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 2682 BP; 767 A; 619 C; 527 G; 769 T; 0 other;
Query Match 2.8%; Score 54; DB 23; Length 2682;
Best Local Similarity 50.8%; Pred. No. 0.0022;
Matches 129; Conservative 0; Mismatches 125; Indels 0; Gaps

QY 114 TGGTGAAGAACCGTAACCTGCTGATGCTGTCAGAGCTTGCAGCCGCAACAATATAGGTCC 173
|||||
Db 1260 TGGTCCCTCATGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 1201
QY 174 TGGAGCCGGTCTTTTGTAGCGGGTTCCTACTTTACAAATATACCGGTGCTTTTACGGTTAC 233
|||||
Db 1200 TGGTCCCTCATGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 1141
QY 234 TGGTCCCTCATGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 293
|||||
Db 1140 AACTGCTGCTGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 1081
QY 294 AGTAACCTGCTGCTGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 353
|||||
Db 1080 AATTGCTGCTGCTGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 1021
QY 354 AGTTAATATTGATG 367
|||||
Db 1020 TGGTCCCTCATGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 1021

RESULT 13
AAA88864
ID AAA88864 standard; DNA; 8201 BP.
XX AC AAA88864;
XX DT 19-FEB-2001 (first entry)
XX DE Human dentin sialophosphoprotein gene.
XX KW Dentin sialophosphoprotein; DSPP; human; SIBLINGS;
XX OS Integrin-binding ligand; diagnosis; osteoporosis; chromosome 4; ds.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT CDS 2387..7896
FT /*tag= a
FT /*note= "contains exons"
FT exon 2387..2437
FT /*tag= b
FT intron 2438..3576
FT /*tag= c
FT exon 3577..3660
FT /*tag= d
FT intron 3661..3793
FT /*tag= e
FT exon 3794..4780
FT /*tag= f
FT intron 4781..5256
FT /*tag= g
FT exon 5257..7896
FT /*tag= h
XX PN WO2000062065-A1..
XX PD 19-OCT-2000.
XX PF 07-APR-2000; 2000WO-US09349.
XX PR 09-APR-1999; 99US-0128468.
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES

XX
SQ Sequence 360 BP; 122 A; 111 C; 81 G; 46 T; 0 other;
Query Match 2.8%; Score 54; DB 23; Length 360;
Best Local Similarity 50.8%; Pred. No. 0.001;
Matches 129; Conservative 0; Mismatches 125; Indels 0; Gaps

QY 114 TGGTGAAGAACCGTAACCTGCTGATGCTGTCAGAGCTTGCAGCCGCAACAATATAGGTCC 173
|||||
Db 260 TGGTCCCTCATGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 201
QY 174 TGGAGCCGGTCTTTTGTAGCGGGTTCCTACTTTACAAATATACCGGTGCTTTTACGGTTAC 233
|||||
Db 200 TGGTCCCTCATGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 141
QY 234 TGGTCCCTCATGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 293
|||||
Db 140 AACTGCTGCTGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 81
QY 294 AGTAACCTGCTGCTGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 353
|||||
Db 80 AATTGCTGCTGCTGCTGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 21
QY 354 AGTTAATATTGATG 367
|||||
Db 20 TGGTCCCTCATGCTGCTGCTGCTGCGCAAGTGTGCGGGTGTGAGATGTTGCTGTC 21

RESULT 12
ABL25080/C
ID ABL25080 standard; DNA; 2682 BP.
XX AC ABL25080;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 26713.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX DT 23-MAR-2001; 2001WO-US09231.
XX PF 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Claim 1; SEQ ID NO 26713; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Fisher LW, Fedarko NS, Young MF;
 XX WPI; 2000-679515/66.
 XX P-PSDB; AAB19772.
 PT Detecting small integrin-binding ligand N-linked glycoproteins for
 PT detection of a tumor or protection against a complement mediated immune
 PT response, comprises detection where Factor H is not an inhibitor
 XX Disclosure; Page 98-100; 110pp; English.
 XX The present sequence is that of DNA encoding human dentin
 CC sialophosphoprotein (DSPP, see AAB19772), a member of the small
 CC integrin binding ligand, N-linked glycoproteins. (SIBLINGS) family.
 CC The invention provides methods and compositions for exploiting the
 CC discovery that members of the SIBLINGS family bind to complement
 CC Factor H, conferring resistance to complement mediated lysis. A
 CC claimed method of conferring protection against a complement
 CC mediated immune response involves providing a reservoir or other
 CC supply in the subject's body so that a SIBLINGS protein can be
 CC dispersed to interfere with complement mediated lysis and
 CC inflammation. This protects cells that are grafted onto foreign
 CC tissue or bone marrow cells introduced into a foreign host. The
 CC SIBLINGS protein can be BSP, OPN, DMPI or DSPP. A method of
 CC detecting a SIBLINGS protein in a sample from a subject suspected
 CC of having abnormal bone turnover, especially osteoporosis, is also
 CC claimed.
 XX SQ Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
 Query Match 2.8%; Score 53.2; DB 21; Length 8201;
 Best Local Similarity 41.1%; Pred. No. 0.0052;
 Matches 379; Conservative 0; Mismatches 543; Indels 0; Gaps 0;
 QY 102 TGAGGTCCTTGGTGTGCAAGAACCGTAACCTGCTGATGTGCGAGAGCTTGACGCGGAAC 161
 DB 7459 TCGTGTCACTGTGCTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7400
 QY 162 AATATAGTCTGGAGCGGGTCTTTGTAGCGGGTCTTACTTTTACAATATACCGGTGC 221
 DB 7399 CACTGTGCTGCTGCTGCTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7340
 QY 222 TTTACGGTTACTGATGCTGACGTAGTGTGCTGCAATGATTTAAATTTTCACG 281
 DB 7339 TGTGCTGCTGCGCACTGCTGCTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCT 7280
 QY 282 AGTCTCTTTTTCAGTAACCTGGTGATATTTTCATTAGTGTTCAGTGGTAGATACGGAGGAGC 341
 DB 7279 TGCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7220
 QY 342 TAATAAATGCTGAGTTAATATTGATGATGGTTTAACTTTAACTTTAAACAGGTACCGGTAC 401
 DB 7219 CACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7160
 QY 402 TGCAGCCTACGGTGCAGAACTCGTGTGTTATTCAGGTGGACAGCTGCTGCTAATAA 461
 DB 7159 TATCACTGCTGTTGCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7100
 QY 462 TACATATATGCTTTTAGGTAAATATAATCTAGTGGAGGGAATCCGGTTTGACTATTGC 521
 DB 7099 TGGTATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7040
 QY 522 TTCAGATCCAGATGATATTAGGACCAATAACGGTTGCGAGGAATATAGATCGAGGAGGTAT 581
 DB 7039 TGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6980
 QY 582 AATACTGACATACAGATGCTGCCAATTAACGGAAACAATAGGTAAATATCCCGCAGC 641
 DB 6979 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6920

642 TCAAAATAGCATTGGAGCAAGTACGCTTCTCTTGAGGGGCGAGTTATTAAAGCCACTAC 701
 6919 TTTCACCTGCTATCACTGCTGCTGCTGCTGCTATCACTGCTGCTGCTGCTGCTGCTATCAC 6860
 702 GACTAAGTTAACAAATGACAGCCGCGGTATTAACACTTTACAAATGCAAAATGCAAGTATTAAC 761
 6859 TGCTGCTCACTATCACTGCTGCTGCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTTCT 6800
 762 AGTCTGCTGTTGATACACACAGCGGGTGCATGATGTAGTGTCTTAAATTTAAACGGAGC 821
 6799 TCGTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 6740
 822 GTTGAGTCAAGTAACCTGGAATATAGGTAATACAAATTCATTAGCGACAAATAGTGTAGG 881
 6739 CCGTGTGCTCACTACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6680
 882 AGCAGGTAGGCTAGCTTAGGGGAGCGGTTATTAAAGCCACTAGCACTAAAGTTAACGAA 941
 6679 CACTGCTATCGCTGCTGCTGCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTAC 6620
 942 TGCAGCGTGGTATTAACACTTACAACTTACAACTTACAACTTACAACTTACAACTTACAACT 1001
 6619 TATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6560
 1002 AGGCGGTGATAATAGTGTGCTC 1023
 6559 TATCACTGCTGCTGCTGCTGCTATC 6538

RESULT 15
 ABO73537
 ID ABO73537 standard; DNA; 8201 BP.
 XX AC ABO73537;
 XX DT 03-OCT-2002 (first entry)
 XX Human dentin sialophosphoprotein precursor (DSPP) gene SEQ ID NO:1.
 DE Human; dentin sialophosphoprotein precursor; dentin sialophosphoprotein;
 KW DSPP; dentinogenesis imperfecta type II; deafness; auditory;
 KW chromosome 4q21; gene; ds.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT promoter 1..6
 FT exon /*tag= a
 FT /*tag= b
 FT /*number= 1
 FT intron 99..2358
 FT /*tag= c
 FT exon /*number= 1
 FT /*tag= d
 FT /*number= 2
 FT CDS 2387..7896
 FT /*tag= e
 FT /*product= "dentin sialophosphoprotein precursor"
 FT /*note= "contains introns"
 FT sig_peptide 2387..2431
 FT /*tag= f
 FT mat_peptide 2432..7893
 FT /*tag= g
 FT /*product= "dentin sialophosphoprotein"
 FT /*note= "contains introns"
 FT mat_peptide 2432..5520
 FT /*tag= h
 FT /*product= "dentin sialoprotein"
 FT /*note= "contains introns"
 FT intron 2438..3576
 FT /*tag= i

FT exon /number= 2
 FT 3577..3660
 FT /*tag= j
 FT /number= 3
 FT 3661..3793
 FT /*tag= k
 FT /number= 3
 FT 3794..4780
 FT /*tag= l
 FT /number= 4
 FT 4781..5256
 FT /*tag= m
 FT /number= 4
 FT 5257..8201
 FT /*tag= n
 FT /number= 5
 FT 5521..7893
 FT /*tag= o
 FT /product= "phosphoryn"
 FT 5596..5604
 FT /*tag= p
 FT /note= "cell binding domain"
 FT 7988..7993
 FT /*tag= q
 FT 8171..8176
 FT /*tag= r
 FT
 FT W0200258722-Al.
 FN
 XX
 XX
 PD 01-AUG-2002.
 XX
 XX
 PF 30-AUG-2001; 2001WO-CN01292.
 XX
 XX
 PR 05-SEP-2000; 2000CN-0125042.
 XX
 XX
 PA (SHAN-) SHANGHAI RES CENT BIOTECHNOLOGY.
 XX
 XX Kong X, Xiao S, Zhao G, Yu C, Hu L;
 XX
 XX WPI; 2002-557897/59.
 DR P-PSDB; ABP51785.
 XX
 XX
 PT
 PT
 PT
 PS
 XX
 XX
 XX

Diagnosis of dentinogenesis imperfecta type III and its accompanying deafness using dentin sialophosphoprotein gene and encoded products - Example 3; Page 23-27; 38pp; Chinese.

The present invention describes a method (M1) for the diagnosis of dentinogenesis imperfecta type II and/or its accompanying deafness comprising determining the dentin sialophosphoprotein (DSPP) gene, its transcript and/or protein of an individual for comparison of their sequences with the normal sequences and judging the individual to have higher risk of suffering from the disease then the normal population. Also described are: (1) treating dentinogenesis imperfecta type III and/or its accompanying deafness by administering a safe and effective dose of normal DSPP and/or DSP protein to patients; (2) drug compositions containing safe doses of DSPP and/or DSP protein; and (3) a reagent kit for detecting dentinogenesis imperfecta type II and/or its accompanying deafness containing primers for specific amplification of DSPP gene or its transcript, or containing probes for binding to the mutation site. The DSPP gene and protein sequences have auditory activity. The method (M1), dentin sialophosphoprotein (DSPP) gene and DSP protein are useful for diagnosing and treating imperfecta type II and/or its accompanying deafness. The DSPP gene is located to chromosome 4q21. The present sequence represents the human DSPP gene from the present invention.

Query Match 2.8%; Score 53.2; DB 24; Length 8201;
 Best Local Similarity 42.3%; Pred. No. 0.0052;
 Matches 423; Conservative 0; Mismatches 568; Indels 9; Gaps 2;
 QY 496 GGAGCGAATGCCGGTTTGACTATTGCTTCAGATCCAGATGATAGGACCAATAACGCTT 555

Db 6227 GTAGTGACAGCAGCAATAGCAGTGTAGAGCAGTGTAGTGTACAGCAGTGTAGTGTAC 6286
 QY 556 GCAGGAATATAGATGGAGGAGGTATATACTACATAACAGATGCTGCCATTAAACGGA 615
 Db 6287 GCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 6346
 QY 616 ACAATAGGTAATACCTAATCCGGCAGCTCAAAATAGCATTGGAGCAAGTACCGTTCTCTT 675
 Db 6347 GCGATAGCAGTGTAGCAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAG 6406
 QY 676 GGAGGCGAGTTTAAAGCCACTACGACTAAAGTTTAAATAAGTGTAGCAGCAGTGTAGTGTAG 735
 Db 6407 GCAGTGTAGCAGCAGCAACAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGTGTAGT 6466
 QY 736 CTTACAAATGCAATGCAAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGTGTAG 795
 Db 6467 ACAGCAGCAGCAGTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGTAGTGTAG 6526
 QY 796 GTAGGTGTCTTTAAATTTAAACGGAGCGTTGAGTGAAGTAACTGGAAATATATAGGTAATACA 855
 Db 6527 GCAACAGCAGTGTAGCAGCAGCAGCAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGT 6586
 QY 856 AATTCATTAGCGACAATTAAGTGTAGGAGCAGGTACGGTACGTTAGGGGGAGCGGTTATT 915
 Db 6587 GCGACAGCAGCAATAGCAGTGTAGCAGCAGTGTAGTGTAGTGTAGCAGCAGCAGTGTAGTGTAG 6646
 QY 916 AAGCCCACTACGACTAAGTTAAACGAATGCAGCGTGGTATTAAACACTTACAAATGCGAGTA 975
 Db 6647 GCAGTGTAGCAGCAGCAGCAGCAGCAGTGTAGCAGTGTAGCAGTGTAGCAGCAGCAGTGTAG 6706
 QY 976 TTAACAGGTGCTGTGTGATAACACACAGCGGCGTGTATATGTAGTGTGTGTAATCTTAAGT 1035
 Db 6707 GAAGTGACAGTAGTAAATAGTAGTGACAGCAGCGATAGCAGTGTAGCAGCAGCAACAGCAGT 6766
 QY 1036 GGAGCATTTAGTCAAGTAACCGGAATATAGGTAATACAAATTCATTAGCCACCAGTAAT 1095
 Db 6767 ACAGCAGTGTAGTGTAGTGTAGCAGCAGTGTAGCAGCAGTGTAGCAGCAGCAGTGTAGCAGCA 6826
 QY 1096 ATAGGAGCAGGTGTGGCTACCTTGGATGGAGCGGTATTAAAGCTACTTACGACTTAAGTTA 1155
 Db 6827 GTGATAGCAGCAACAGCAGTGTAGTGTAGCAGCAGTGTAGCAGCAGCAGTGTAGCAGCA 6886
 QY 1156 ACAGATGCGGTGAGTGTATTAATTTACGAATCCTCTGTAGTAGT-----AACCGAGCA 1209
 Db 6887 GTGATAGCAGCAACAGCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 6946
 QY 1210 ATAGATAATACCGGTAAATGCCAATAAAGGTGTGGTAAATCTTTACCGGAGCAAGTACGTA 1269
 Db 6947 GCAGTGTAGCAGCAACAGCAGTGTAGTGTAGCAGCAGTGTAGTGTAGCAGCAGTGTAGTGTAGT 7006
 QY 1270 ACCGATAATATAGTGTAAACGGCAGTATTAGCAGAGGTAAAGCGGTAGGAGCAGGTTTGTCTG 1329
 Db 7007 ACAGCAATATAGCGGTGACAGCAGCAACAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGTGTAG 7066
 QY 1330 CAATACAGCGCGAGTGTAAAGCAATGCAATTAACCTTAACCGGATATGCGTGTAGTGTAGTGTAGT 1389
 Db 7067 ACAGCAGTGTAGCAGCAACAGCAGCAGTGTAGTGTAGCAGCAGTGTAGTGTAGTGTAGTGTAGTGTAG 7126
 QY 1390 GT---AACATTTACCGGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1446
 Db 7127 GTGACAGCAGTGTAGCAGCAACAGCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 7186
 QY 1447 GCAACAGTGTAGTGTAGCAGCAATTAACATTACGAGCGG 1486
 Db 7187 GCAGTGTAGTGTAGTGTAGCAGCAGCAACAGCAGTGTAGCAGCAG 7226

Search completed: April 17, 2003, 09:22:06
 Job time : 1006 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 12:04:21 ; Search time 193 Seconds
(without alignments)

3050.878 Million cell updates/sec

Title: US-09-800-065-1

Perfect score: 1920

Sequence: 1 atggcggaatttctctaaa.....cgttgagctaaacaatag 1920

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	60.4	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
2	57.6	3.0	3666	2 US-08-682-517-13	Sequence 13, Appl
3	57.6	3.0	3666	2 US-08-682-517-14	Sequence 14, Appl
4	57.6	3.0	4197	2 US-08-682-517-7	Sequence 7, Appl
5	57.6	3.0	4197	2 US-08-682-517-8	Sequence 8, Appl
6	49.6	2.6	2793	1 US-08-209-747-1	Sequence 1, Appl
7	49.6	2.6	2793	1 US-08-458-298-1	Sequence 1, Appl
8	46.8	2.4	5361	4 US-08-973-462-2	Sequence 2, Appl
9	46.8	2.4	6152	4 US-08-973-462-1	Sequence 1, Appl
C 10	46.8	2.4	543	6 5273901-6	Patent No. 5273901
C 11	45.8	2.4	397	3 US-09-253-691-3	Sequence 3, Appl
12	45.8	2.4	5319	1 US-08-169-927-1	Sequence 1, Appl
13	43.6	2.3	1891	4 US-08-973-462-3	Sequence 3, Appl
14	43.4	2.3	789	3 US-08-483-857-3	Sequence 3, Appl
15	43.4	2.3	1800	3 US-08-483-857-1	Sequence 1, Appl
16	43.2	2.3	1173	4 US-09-134-001C-281	Sequence 281, App
17	43.2	2.3	2169	4 US-09-434-408-3	Sequence 3, Appl
C 18	42.6	2.2	234	1 US-08-469-802B-3	Sequence 3, Appl
C 19	42.6	2.2	234	2 US-08-267-803B-3	Sequence 3, Appl
C 20	41.6	2.2	203	4 US-09-043-303-7	Sequence 7, Appl
C 21	41.6	2.2	477	4 US-09-135-994-1	Sequence 1, Appl
C 22	41.6	2.2	3168	4 US-09-165-239A-3	Sequence 3, Appl
C 23	41.4	2.2	740	4 US-09-451-117-1	Sequence 1, Appl
24	41.4	2.2	2472	4 US-09-134-001C-1244	Sequence 1244, App
C 25	40.8	2.1	1267	4 US-09-071-035-483	Sequence 483, App
C 26	40.8	2.1	1356	4 US-09-071-035-481	Sequence 481, App
27	40.4	2.1	2830	2 US-09-010-928B-1	Sequence 1, Appl

C 28	40.4	2.1	8585	1 US-08-030-096-3	Sequence 3, Appl
29	39.4	2.1	10851	2 US-08-286-819A-16	Sequence 16, Appl
30	39.4	2.1	10851	3 US-08-980-357-16	Sequence 16, Appl
31	39.2	2.0	789	1 US-08-436-748-4	Sequence 4, Appl
32	39.2	2.0	1800	1 US-08-436-748-1	Sequence 1, Appl
C 33	39	2.0	533	6 5482709-5	Patent No. 5482709
C 34	39	2.0	4248	3 US-08-678-614-1	Sequence 1, Appl
35	39	2.0	4287	1 US-08-038-682-7	Sequence 7, Appl
36	39	2.0	4287	1 US-08-302-832-7	Sequence 7, Appl
37	39	2.0	4287	2 US-08-530-198-7	Sequence 7, Appl
38	39	2.0	4287	2 US-08-469-880-7	Sequence 7, Appl
39	39	2.0	4287	2 US-08-728-470-7	Sequence 7, Appl
40	39	2.0	4287	4 US-08-719-641-7	Sequence 7, Appl
41	39	2.0	4702	1 US-08-038-682-8	Sequence 8, Appl
42	39	2.0	4702	1 US-08-302-832-8	Sequence 8, Appl
43	39	2.0	4702	2 US-08-530-198-8	Sequence 8, Appl
44	39	2.0	4702	2 US-08-469-880-8	Sequence 8, Appl
45	39	2.0	4702	2 US-08-728-470-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match

3.1%; Score 60.4; DB 1; Length 7218;

```
Best Local Similarity 4.8%; Pred. No. 4.4e-06;
Matches 19; Conservative 221; Mismatches 152; Indels 0; Gaps 0;

QY 1148 CTAAGTAAACAGATGATCGGTACGATTAATTAATTTACGAATCCTGCTAGTAAACCGGAG 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1471 CTATCAAGTAGTTAAACAGATAGAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRR 1412

QY 1208 CAATAGATAATACCGGTAAATGCCATAAAGGTGGTAAATCTTTACCGGAGCAAGTACGG 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1411 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352

QY 1268 TAAACGATAATAGTTAAACAGCGCAGTATTAGCAGAGGTAAAGCTAGGACAGGTTTGC 1327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1351 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1292

QY 1328 TCGAATAACAAGGGGAGTAGTAAAGCGAATGCAATAAATTAACGGATAATAGCGTCAG 1387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1291 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1232

QY 1388 TAGTAACAATTTACCGGTGATAGTACGGTAACAGGTAGTATAGGTGGTACAGAACTCTTCG 1447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172

QY 1448 CAACAGTCAATATAGGACGAGCAATAACATTACGAGCGGAGGAGCCTAGCTCGGAATA 1507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112

QY 1508 ATATAGATTTTGGAGTCTGCGAGTAAATTTAGAG 1539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1111 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080

RESULT 2
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-682-517-13

Query Match 3.0%; Score 57.6; DB 2; Length 3666;
Best Local Similarity 47.0%; Pred. No. 1.8e-05;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

QY 116 CTGCAAGAACCGTAACTGCTGATGCTGAGCTTGCAGAGCTTGCAGCCGGAACAAATATAGTCTGT 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2735 CTTTAGGAATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2794

QY 176 GAGCGGTGCTTTGTAGCGGTCTTCTTACAAATATACGGTCTTTACGGTTACTG 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2795 CACTGTTTCATTAAG-----ATAGTCAATAATTCATTATCTTCTTACATAGTTG 2848

QY 236 ATGCTGACGTAAGTGTTCGTGCATAGATTAATAATTTTGCAGCAGGCTTTTTTTCAG 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2849 AAACGTGCTGCTAATACAGGTGATTTCTTACAACTGTTCAGCTGTTCAAGCTGTTACATCTTCT 2908

Best Local Similarity 4.8%; Pred. No. 4.4e-06;
Matches 19; Conservative 221; Mismatches 152; Indels 0; Gaps 0;

QY 1148 CTAAGTAAACAGATGATCGGTACGATTAATTAATTTACGAATCCTGCTAGTAAACCGGAG 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1471 CTATCAAGTAGTTAAACAGATAGAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRR 1412

QY 1208 CAATAGATAATACCGGTAAATGCCATAAAGGTGGTAAATCTTTACCGGAGCAAGTACGG 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1411 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352

QY 1268 TAAACGATAATAGTTAAACAGCGCAGTATTAGCAGAGGTAAAGCTAGGACAGGTTTGC 1327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1351 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1292

QY 1328 TCGAATAACAAGGGGAGTAGTAAAGCGAATGCAATAAATTAACGGATAATAGCGTCAG 1387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1291 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1232

QY 1388 TAGTAACAATTTACCGGTGATAGTACGGTAACAGGTAGTATAGGTGGTACAGAACTCTTCG 1447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172

QY 1448 CAACAGTCAATATAGGACGAGCAATAACATTACGAGCGGAGGAGCCTAGCTCGGAATA 1507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112

QY 1508 ATATAGATTTTGGAGTCTGCGAGTAAATTTAGAG 1539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1111 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080

RESULT 3
US-08-682-517-14
; Sequence 14, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
; US-08-682-517-14

Query Match 3.0%; Score 57.6; DB 2; Length 3666;
Best Local Similarity 47.0%; Pred. No. 1.8e-05;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

QY 116 CTGCAAGAACCGTAACTGCTGATGCTGAGCTTGCAGAGCTTGCAGCCGGAACAAATATAGTCTGT 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2735 CTTTAGGAATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2794

QY 176 GAGCGGTGCTTTGTAGCGGTCTTCTTACAAATATACGGTCTTTACGGTTACTG 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2795 CACTGTTTCATTAAG-----ATAGTCAATAATTCATTATCTTCTTACATAGTTG 2848

QY 236 ATGCTGACGTAAGTGTTCGTGCATAGATTAATAATTTTGCAGCAGGCTTTTTTTCAG 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2849 AAACGTGCTGCTAATACAGGTGATTTCTTACAACTGTTCAGCTGTTCAAGCTGTTACATCTTCT 2908
```

Db 2795 CAACGTGTTTCATTAAG-----ATAGTGCATAAATTAATTCATATCTCTTACATAGTTG 2848
QY 236 ATGCTGAGCTAGTCTCGTGCATTAGATTTAAATAATTTTGCACAGGTCTTTTTCAG 295
Db 2849 AAACGGTCTAATACAGGTGATTTTGTACAACTGTTCAAGCTGGTCAATATCTTCTT 2908
QY 296 TAACTGGTGATATTTCAATAGTTCAGTGGTAGATACGGGAGGAGCTAATAAATTCGAG 355
Db 2909 TAACTGGTGATATTTCAATAGTTCAGTGGTAGATACGGGAGGAGCTAATAAATTCGAG 2968
QY 356 TTAATATTGATGATGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT 415
Db 2969 AAAATATTACTGCTAGCGTAACATTAAGAAAC-----TACTGGAGCAATTTACTTCTGATA 3025
QY 416 CAAATCCTCGGTTGTTTCCAAAGTGGACAAAGCTGCTGCTAATAATACATATATCTGCTT 475
Db 3026 CATTTACAAAGGTGATTAACCATGAGCAAGCTGACAGAGCTGAATATATCTTCTAATCAA 3085
QY 476 TAGGTAATATACTCTAGTGGAGGGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG 535
Db 3086 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACITTTAAATATTGATAATGCTG 3145
QY 536 TATTAGGACCAATACGCTTGACAGAAATATAGATGGAGGAGGTATATAAATGACAATA 595
Db 3146 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGAGTGCACAAAGT---GTAGCTGATGCTA 3202
QY 596 CAGATGCTGCCAATTAACGGAACAATAGGTAACTAAATCCGCGAGCTCAAAATAGCAATG 655
Db 3203 TCATGCTACATTTGCGAGTACTGCACACTGTTCTTGGAGACAAGTAGTTTAAATCAAG 3262
QY 656 GAGCAAGTACGCTTCTTCTTGGAGGGGCAAGTTATTAAAGCCACTACGACTAAGTTAAACAA 715
Db 3263 CTACAACAGGTGTTGTTCTTGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3322
QY 716 ATGCAG 721
Db 3323 ATGCAG 3328

RESULT 4

US-08-682-517-7

; Sequence 7, Application US/08682517

; Patent No. 5874267

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Expression of surface layer proteins

; NUMBER OF SEQUENCES: 25

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/682,517

; FILING DATE:

; CLASSIFICATION:

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4197 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-682-517-7

Query Match

Best Local Similarity 3.0%; Score 57.6; DB 2; Length 4197;

Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

QY 116 CTGCAAGAACCGTAACCTGCTGATGGTGCAGAGCTTCAGCCGGAACAATAATAGGTCCTG 175
Db 2919 CTTAGGAATTCATTAGCTGATGCAGATCTTAATGAATGCAACAACCTGTTGATACGTG 2978

QY 176 GAGCGGTGCTTTTGTAGCGGTTCTACTTTACAATAATACCGGTGCTTTTACGGTTACTG 235
Db 2979 CAACGTGTTTCATTAAG-----ATAGTGCATAAATTAATTCATATCTCTTACATAGTTG 3032
QY 236 ATGCTGAGCTAGTCTCGTGCATTAGATTTAAATAATTTTGCACAGGTCTTTTTCAG 295
Db 3033 AAACGGTCTAATACAGGTGATTTTGTACAACTGTTCAAGCTGGTCAATATCTTCTT 3092
QY 296 TAACTGGTGATATTTCAATAGTTCAGTGGTAGATACGGGAGGAGCTAATAAATTCGAG 355
Db 3093 TAACTGGTGATATTTCAATAGTTCAGTGGTAGATACGGGAGGAGCTAATAAATTCGAG 3152
QY 356 TTAATATTGATGATGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT 415
Db 3153 AAAATATTACTGCTAGCGTAACATTAAGAAAC-----TACTGGAGCAATTTACTTCTGATA 3209
QY 416 CAAATCCTCGGTTGTTTCCAAAGTGGACAAAGCTGCTGCTAATAATACATATATCTTCTT 475
Db 3210 CATTTACAAAGGTGATTAACCATCAGCAGCTACAGCAGCTGAATATATCTTCTAATCAA 3269
QY 476 TAGGTAATATACTCTAGTGGAGGGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG 535
Db 3270 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACITTTAAATATTGATAATGCTG 3329
QY 536 TATTAGGACCAATACGCTTGACAGAAATATAGATGGAGGAGGTATATAAATGACAATA 595
Db 3330 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGAGTGCACAAAGT---GTAGCTGATGCTA 3386
QY 596 CAGATGCTGCCAATTAACGGAACAATAGGTAACTAAATCCGCGAGCTCAAAATAGCAATG 655
Db 3387 TCAATGCTACATTTGCGAGTACTGCACACTGTTTCTTGGAGACAAGTAGTTTAAATCAAG 3446
QY 656 GAGCAAGTACGCTTCTTCTTGGAGGGGCAAGTTATTAAAGCCACTACGACTAAGTTAAACAA 715
Db 3447 CTACAACAGGTGTTGTTCTGGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3506
QY 716 ATGCAG 721
Db 3507 ATGCAG 3512

RESULT 5

US-08-682-517-8

; Sequence 8, Application US/08682517

; Patent No. 5874267

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Expression of surface layer proteins

; NUMBER OF SEQUENCES: 25

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/682,517

; FILING DATE:

; CLASSIFICATION:

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4197 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: Bacillus sphaericus

; INDIVIDUAL ISOLATE: P-1

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 95..3850

; FEATURE:

; NAME/KEY: mat_peptide

```
; LOCATION: 185...3950
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95...184
US-08-682-517-8

Query Match          3.0%; Score 57.6; DB 2; Length 4197;
Best Local Similarity 47.0%; Pred. No. 1.9e-05;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

QY 116 CTGCAAGAACCCGTAAGTCTGTGTCAGAGCTTGCAGCGGAGCAAAATATAGCTCCTG 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2919 CTTAGGAATTCATTAGCTAGTGCAGATCTTAATGTAAGTGCACAACTGTGATACG 2978

QY 176 GAGCGGTGCTTTTGTAGCGGTCTTACTTTACAATATACCGGTCTTTTACGGTTACTG 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2979 CAACGTGTTTCATTAAGG-----ATAGTGCAAATAATTCATTCTCTTACATTAGTG 3032

QY 236 ATGCTGAGCTAGTGTCTGTCATTAGATTAAATAATTTGCAGCAGGTCTTTTTCAG 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3033 AAACGTGGTCTAATACAGGTGTATTGTACAACCTGTTCAAGCTGGTACATTATCTTCT 3092

QY 296 TAACGTGTCATATTCATTAGTTTCAGTGTAGATACGGGAGGAGCTAATAAAGCTTCAG 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3093 TAACGTGCTGGTACATTACAGTTACTTATGAGATGCTTAAAGTCTCCAGGTGTGCTG 3152

QY 356 TTAATATTGATGAGGTGTTTAACTTTAACTTTAACAGGTACCGGTACTGCAGCCTACGGTG 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3153 AAAATATTACTGCTAGCGTAACATTAAAGAAAC---TACTGGAGCAATTACTCTGATA 3209

QY 416 CAAATCCCTGCTGTTATTCCAAAGGTGCACAGCTGCTGCTTAATAATACATATAGTGT 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3210 CATTTACACAGGTGTATTCCATCAGCAGCTACAGCAGCTGAATATACTTCTAAATCAA 3269

QY 476 TAGGTAATATAACTCTAGTGGAGCGAATGCCGGTTTCACATTTGCTTCAGATCCAGATG 535
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3270 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACTTTAAATATGATATGCTG 3329

QY 536 TATTAGGACCAATACGGTTCGAGAAATATAGATGGAGGAGGTATATAACTGACAATA 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3330 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGAGTGCACAAAGGT---GTAGCTGATGCTA 3386

QY 596 CAGATGCTGCCATTAACGGAACAATAGTAACTACTAATCCGGCAGCTCAATAAGCATG 655
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3387 TCATGCTACATTTGCAGGTACTGCACTGTTCTTGAGACAAAGTAGTTATTAAATCAG 3446

QY 656 GAGCAAGTACGCTTCTCTTTGGAGGGGAGTTATTAAAGCCACCTACGACTAAGTTAAACA 715
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3447 CTACAACAGGTGTTGGTTCTGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTA 3506

QY 716 ATGCAG 721
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3507 ATGCAG 3512

RESULT 6
US-08-209-747-1
; Sequence 1, Application US/08209747
; Patent No. 573771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: GDNAS Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2793 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Nephila clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 183..2675
; OTHER INFORMATION: /product= "N. clavipes minor
; ampullate silk protein"
; US-08-209-747-1
```

```
Query Match          2.6%; Score 49.6; DB 1; Length 2793;
Best Local Similarity 41.5%; Pred. No. 0.0019;
Matches 157; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 95 CCGTAGTGGAGTCTTGTGCTGCAAGAACCGTAACCTGTGTGTCGAGAGCTTGCGAG 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 CTGCTGTGGAGCTGGAGCCGCTGCAGTGCAGGAGGAGCTGGAAGCTACCGTGGTC 951

QY 155 CCGCAACAATATAGTCTGGAGCCCGGTGCTTTTGTAGCGGTCTACTTTACAATA 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 952 AAGGATACGTCGCGGAGCAGAGCTGTGCTGCTGCGAGTGCAGGTGCTGGATATGGTGGAGAG 274

QY 215 CCGTGTCTTTTACGTTTACTGCTCAGCTAAGTGTTCGTGCTGCTGCAANNNNNNNNNN 1011
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1012 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1071

QY 275 TTGCAGCAGTCTTTTTCAGTAACCTAGTGTGATATTTTCATTAGGTTCAGTGTAGATACGG 334
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1072 CGGATATGTCGCGGAGCAGGAGCTGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1131

QY 335 GAGGAGCTAATAAATTCAGTAAATTAATGATGATGTTTAACTTAACTTTAACAGGTA 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 GTGCTGTGTTTACGGTAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191

QY 395 CCGTACTGCAGCTACGCTGCAAAATCCTGCTGTTGTTTATTTCCAAAGTGGACAGCTGCTG 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1192 CAGGAGCTGGAGCTAGGTTGGTCAAGTGGATACGCTGCGGAGCAGGAGTGTGCTGCTGCG 1251

QY 455 CTAATAATACATATACG 472
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1252 CTGCTGAGCAGGAGCTG 1269

RESULT 7
US-08-458-298-1
; Sequence 1, Application US/08458298
; Patent No. 5756677
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
```


Qy	615	AACATAGGTAACTAATAATCCGGCAGCTCAAAATAAGCATTTGAGCAAGTAGCGTTCCTCT	674
Db	1389	AGCTGAATAATGTTTGAAAGAAAGTGTAGCTGAAAATCTTGAAGAAAGTGTAGCTGAAAATGT	1448
Qy	675	TGAGGGGCAGTTATTAAAGCCACTACGACTAAGTTTACAATAATCGACGCCGGTATTAAAC	734
Db	1449	TGAAGAAGGTGTAGCTGAAAAATGTTTGAAGAAAGTGTAGCTGAAAATCTTGAAGAAAGTGT	1508
Qy	735	ACTTACAAATGCCAAATCGAGTATTAAACAGGTGCTGCTTCAATACACCACACGCGGTGATGA	794
Db	1509	AGCTGAATAATGTTGAAGAAAGTGTAGCTGAAAATGTTTGAGAAATCGTAGCTCCAACGT	1568
Qy	795	TGTAGTGTGCTTTAAATTTAAACGGACGGTTGAGTCAAGTAACCTGGAATATATAGTAATAC	854
Db	1569	TGAAGAAATCGTAGCTCCAACCTGTTTGAAGAAATTGTAGCTCCAAGTGTGTAGNAAGTGT	1628
Qy	855	AAATTCATTAGCCACAAATAAGTGTGTAGGAGCAGGTACGGCTACGTTAGGGGAGCGGTTAT	914
Db	1629	GGCTCCAAAGTGTGTGAAGAAAGTGTAGAAGAAATGTTTGAAGAAAGTGTAGCTGAAAATGT	1688
Qy	915	TAAAGCCACTACGCACTTAAGTTAACGAATCGACGCTCGGTATTAAACACTTACAATGCAGT	974
Db	1689	TGAAGAAAGTGTAGCTGA---AAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAG	1745
Qy	975	ATTAACAGGTGCTGTTGTATPAACACACACAGCGCGTGATAATGTAGGTGTCGTAATCTAAG	1034
Db	1746	TGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACCTGTTGAAGAAATCGTAGCTCCAAC	1805
Qy	1035	TGGAGCATTCAGTCAAGTAAACCGGAATAATAGTTAATACAATAATTCATTAACGACGATAMA	1094
Db	1806	TGTTGAAGAAATGTAGCTTCCAAGTGTGTGAAGAGTGTGGCTCCAAGTGTGAAGAAAG	1865
Qy	1095	TATAGGACAGGTGTGGCTACCTTCGGATGGAGCGGTTATTAAAGCTACTACGACTAAGTT	1154
Db	1866	TGTGAAGAAATGTTGAAGAAAGTGTASCTGAAAATGTTGAAGAAAGTGTAGCTGAAAAT	1925
Qy	1155	AACAGATGATGCGTCAGTATTAAATATPACGAATCCTGTAGTAGTAAACCGGACCAATAGA	1214
Db	1926	TGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACCTGTTCAAGAAAT	1985
Qy	1215	TAATACCGGTAAATGCCAATAAGGTGTGGTAACTCTTTACCGGACGACGTACGTAACCGA	1274
Db	1986	CGTAGCTCCAACCTGTTGAAGAAATGTTAGTCTCAAGTGTGTGAAGAAAGTGTGGCTCCAAG	2045
Qy	1275	TAAATAGTATAACCGGACGATTATAGCAGAGGTAAAGCTAGGTAGGACAGGTTTGCTGCAAAT	1334
Db	2046	TGTTGAAGAAAGTGTAGAAGAAATGTTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAG	2105
Qy	1335	ACNAGCGGAGTAGTAAAGCGAATGCA	1362
Db	2106	TGTAGCTGAAAATGTTGAAGAAAGTGA	2133

RESULT 9

```

US-08-973-452-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUERSTIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA

```

; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 2.4%; Score 46.8; DB 4; Length 6152;
Best Local Similarity 40.9%;
Pred. NO. 0.013;
Matches 527; Conservative 0; Mismatches 752; Indels 9

QY	75	AACCGCAGCGATAAATGCTTAACCGGTAGTGGAGTGCCTTGGTGTGCAAGAACCGTAACCTGC	134
Db			
Db	1099	AATCGTAGCTCCAAAGTGTTGTAGAAAGTGTGCCTCAAGTGTGAAGAAAGTGTAGAAGA	1155
QY			
Db	135	TGATGCTGACAGCTTGCAGCGGAACAATATAGTCTCTGCAGCGGTGCTGTTTGTAGC	194
Db			
Db	1159	AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGA	1218
QY	195	GGTTCTACTTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACGTAAGTGTTCG	254
Db			
Db	1219	AAGTGTAGCTGAAATGTTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATGTAAGTACCTC	1278
QY	255	TGCATTAGATTTAAATAAATTTTGCACGAGTCTTTTTCAGTAACTGGTGATATTTCAAT	314
Db			
Db	1279	AACTGTTGAAGAAA-----TCGTAGCTCCAACTGTTGAAGAAATTTAGCTCCAAAGTGT	1332
QY	315	AGTTTCAGTGGTAGATACGGGAGGAGCTAATAAAGCTTGCAGTGAATATTTGATGATGGT	374
Db			
Db	1333	TGTAGAAAGTGTGGCTCCAAGTGTGAAGAAAGTGTGAAGAAATGTTGAAGAAAGTGT	1399
QY	375	AACCTTAACTTTAACAGGTTACCGGTACTGACGCCCTACGGTGCAATCTCTGGTGTATT	434
Db			
Db	1393	AGCTGAAATGTTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATG	1452
QY	435	CCAAGTGTGACAAAGTGTCTTAATAATACATATACTGCTTAGGTAAATATAAAGTCTAGG	494
Db			
Db	1453	TGAAGAAAGTGTAGCTGAAATGTTGMAAGAAAGTGTAGCTGAAATGTTGAAGAAATG	1512
QY	495	TGAGCGAATGCGGTTTGTGACTATTGCTTTCAGATCCAGATGTATTAGACCAATTAAGCGT	554
Db			
Db	1513	AGCTCAAGCTGTTGAAGAAATGCTAGCTCCAAGTGTGAAGAAATGTTAGCTCCAAGTGT	1572
QY	555	TGCAGGAAATATAGATGGAGGAGGTATAATACTCACAAATACAGATGCTGCCATTAAAGG	614
Db			
Db	1573	TGTAGAAAGTGTGGCTCCAAGTGTGAAGAAAGTGTAGACAAATGTTGAAGAAAGTGT	1632
QY	615	ACAATAGTATACTAATATCCGGCAGCTCAAAATAAGCAATTTGGAGCAAGTACGCTTCTCT	674
Db			
Db	1633	AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATGTTAGCTGAAATG	1692
QY	675	TGAGGGCGAGTTATTAAAGCCACTACGACTAAGTTTAACAATGCAGCGCGGTATTAAAC	734
Db			
Db	1693	TGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGT	1752
QY	735	ACTTACAAATGCAATTCAGTATTAAACAGGTGCTGTTGATAACCAACACAGCGCGGTATGA	794
Db			
Db	1753	AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATGTTAGCTCCAAGTGT	1812
QY	795	TGTAGTGTCTTTAAATTTTAAACGGAGCGTGTGAGTCAAGTAAGTCTGGAATATAGTAAATAC	854
Db			
Db	1813	TGAAGAAATGCTAGCTCCAACTGTTGAAGAAATTTAGTCTCCAAGTGTGTGAGAAAGTGT	1872
QY	855	AAATTCATTAGCGCAATTAAGTGTAGGAGCAGGTACGGCTACGTTAGGGGAGCGGTATAT	914
Db			
Db	1873	GGCTCCAAGTGTGAAGAAAGTGTAGAAGAAATGTTGAAGAAAGTGTAGCTGAAATG	1932
QY	915	TAAAGCCACTACGACTAAGTTTAAAGAAATGACAGGTGCGGTATTACACTTTACAATGCAAGT	974
Db			
Db	1933	TGAAGAAAGTGTAGCTGA---AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAG	1989
QY	975	ATTAAACAGTGTGTTGATAACACCACAGCGGTGATAATGTAGGTGCTGTAATCTTAAG	1034
Db			
Db	1990	TGTAGCTGAAATGTTGAAGAAATGCTAGCTCCAAGTGTGAAGAAATGTAAGTCTCCAACT	2049
QY	1035	TGAGCAATGAGTCAAGTAAACCGGAATATAGGTAATAACAATTTCAATTAAGCAGGATAAA	1094
Db			

Db 2050 TGTGTAAGAAATTTAGCTCCCAAGTTTCTAGAAAGTTGGCTCCCAAGTTTGAAGAAG 2109
QY 1095 TATAGACAGAGTGTGGCTACCTTGGATGGAGCGGTTATTAAAGCTACTACGACTAAGTT 1154
Db 2110 TGTAGAAGAAATTTGTAAGAAAGTTAGCTGAAAATGTTGAAGAAAGTTAGCTGAAA 2169
QY 1155 AACAGATGATCGTGCAGTATTATATTTACGAATCCTGTAGTAGTACCGGGAGCAATAGA 1214
Db 2170 TGTGAAGAAAGTTAGCTGAAAATGTTGAAGAAATCTAGCTCCCACTGTTGAAGAAT 2229
QY 1215 TAATACCGGTAAATGCAATAAGAGTGTGTAATCTTTTACCGGAGCAAGTACGGTAACCGA 1274
Db 2230 CGTAGCTCAACTGTTGAAGAAATTTAGCTCCCAAGTGTGAGAAAGTTGGCTCCCAAG 2289
QY 1275 TAATAGTGTACACGGCAGTATTAGCAGAGTAAAGCAATGCA 1362
Db 2290 TGTGAAGAAAGTTAGAGAAATGTTGAAGAAAGTTAGCTGAAAATGTTGAAGAAG 2349
QY 1335 ACAAGCGGAGTAGTAAAGCGCAATGCA 1362
Db 2350 TGTAGCTGAAAATGTTGAAGAAAGTTGA 2377

RESULT 10
5273901-6/c
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOOROZITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO:6:
; LENGTH:543
5273901-6

Query Match 2.4%; Score 46.4; DB 6; Length 543;
Best Local Similarity 53.3%; Pred. No. 0.0067;
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 110 TTGCTCTCGCAAGAACCGTAACCTGCTGATGTTGTCAGAGCTGCAGCGCAACAAATATAG 169
Db 244 TTGCTCTCTGCTGCTGCGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 185
QY 170 GTCTCTGAGCGGCTGTTTGTAGCGGGTCTACTTTTACAATATACCGTGTCTTTTACGG 229
Db 184 TTGCTGTAGCTGCTGCTGTTGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 125
QY 230 TTAGTATGCTGACGTAAGTGTTCGGCAATAGATTTAAATATTTTTCAGCAGGCTCTTT 289
Db 124 CTGCTGTGCTGCTGCAAGTGTGCGGCTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTC 65
QY 290 TTTC 293
Db 64 TCTC 61

RESULT 11
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN

; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 2.4%; Score 45.8; DB 3; Length 397;
Best Local Similarity 48.0%; Pred. No. 0.0085;
Matches 131; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 76 ACCGAGCGGATAATGCTAACCGGTAGTGGAGTCCCTTGGTGTGCAAGAACCGTAACGTCT 135
Db 377 AACTGTGCTGATAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
QY 136 GATGTTGACAGCTTTCAGCGGAAATAATATAGTCTTCGAGCCCGGCTCTTTTGTAGCG 195
Db 317 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258
QY 196 GGTTCCTACATATATACCGGCTGCTTTTACGGTTACTGATGCTGACGTAAGTGTCT 255
Db 257 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 198
QY 256 GCATTAGATTTAAATAATTTTCAGCAGGCTCTTTTTCAGTAACCTGCTGATATTTCAATTA 315
Db 197 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138
QY 316 GGTTCAGTGTGATACAGGAGGAGCTAATAAA 348
Db 137 GTTGCTGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 105

RESULT 12
US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:

```

? NAME: Spevack, A. David
? REGISTRATION NUMBER: 24,743
? REFERENCE/DOCKET NUMBER: 75,976
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 295-6759
? TELEFAX: (301) 295-1022
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5319 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: circular
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE: Internal
? ORIGINAL SOURCE:
? ORGANISM: Rickettsia prowazekii
? STRAIN: Breinl
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 340..345
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 363..368
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 391..5226
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 379..386
? FEATURE:
? NAME/KEY: stem_loop
? LOCATION: 5270..5306
? PUBLICATION INFORMATION:
? AUTHORS: Carl, M.
? AUTHORS: Dobson, M. E.
? AUTHORS: Ching, W. M.
? AUTHORS: Dasch, G. A.
? TITLE: Characterization of the
? TITLE: protective S-layer protein
? TITLE: prowazekii; presence of a
? TITLE: homolog in rickettsia typhi
? JOURNAL: Proc. Natl. Acad. Sci.
? DATE: 1990
? RELEVANT RESIDUES IN SEQ ID NO:
? US-08-1699-927-1

```

Query Match	2.48;	Score 45.6;	DB 1;	Length 5319;
Best Local Similarity	49.6%;	pred. No. 0.026;		
Matches 117;	Conservative 0;	Mismatches 119;	Indels 0;	Gaps 0;
QY 1488	AGGAAGCCTAGCTGCGAATAATATAGATATTTGGAGCTGCGAGTAATTTACGAGTTTAACGG	1547		
Db	1917	AGCTAACGATGCTTCAAAATATATAGCACTCGATGGCGCAAAATATATCGGGGCTAATGT	1976	
QY 1548	TCCTGCGCGTAAGAATTTATTAACCTTAATCGAACTATAGCAAAACGGTAATAATGCTACACT	1607		
Db	1977	TGTTGTGCGCAATTCATTTTCCAAAGCTAACGGTGCTACTATTAAATTAACAAATACTCAAA	2036	
QY 1608	TAATATTAATGCTGCTGGTACAGTGATTGCAAAATGATTTAGTAGTAGGCACAGTTGTCACA	1667		
Db	2037	TAATATTGTAGTTAAATTTTCATTTAGATATAACTACTGATAAACACAGTGTGTTGATGC	2096	
QY 1668	AATTAACAATTCAAAATAAATGAATTTTTTGTAAATAATGCTAAGAACGCGCTGATGTTG	1723		
Db	2097	AAGTAGTTTAAACAAATAATCAACTTTTAACCTATTATATGTTAGTAGTATCGGTACTGTTG	2152	

RESULT 13
US-08-973-462-3
; Sequence 3, Application US/08973462B
; Patent No. 6191270

```

; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1891
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1891)
; OS-08-973-462-3

```

Query Match	2.3%;	Score 43.6;	DB 4;	Length 1891;
Best Local Similarity	43.9%;	Pred. No. 0.058;		
Matches 282;	Conservative 0;	Mismatches 354;	Indels 6;	Gaps 2;

QY	75	AACGCGACGATATATGCTAACCCGGTACTGGAGTCTTTGGTGTGCTGCAAGAACCGTAAC	TGC	134
Db				
QY	505	ATCTGTAGTCCCACTGTTCAAGAAATTTGAGTCCCAAGTGTGTGAAAGTGTGGGTCC	564	
Db				
QY	135	TGATGGTGCAGAGCTTGCACGCCGAACAAATATAGTCTCGGACCGGTGCTTTTGTAGC	194	
Db				
QY	565	AAGTGTGTAAGAAAGTGTAGTCCCAAGTGTGGAAGAAAGTGTAGCTCAAAATGTTGAAG	624	
Db				
QY	195	GGTCTCTACTTTACAATATACCCGGTGTCTTTACGGTTACTTCAGTCTCAGCTAAGTGTTCG	254	
Db				
QY	625	AAGTGTAGCT---GAAATGTTGAAGAAATCGTAGTCTCCCAAGTGTGTGAAGAAAGTGTAGC	681	
Db				
QY	255	TGCATTAGATTTAAATAATTTTGCACAGGTCTTTTTTCAGTAACTGGTGATATTTTCATT	314	
Db				
QY	682	TGAAATGTTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTCAAAATGTTGA	741	
Db				
QY	315	AGGTTCAAGTGTAGATACGGGAGGAGCTATAAACITTCGAGTTTAATATTGATGATGTTT	374	
Db				
QY	742	AGAAAGTGTAGCTGAAATGTTGAAG---AAGTGTAGCTGAAATGTTGAAGAAATCGT	798	
Db				
QY	375	AACCTTTAACTTTAACAGGTACCCGGTACTGCAGGCTACGGTGCAAATCCTCGTGTATT	434	
Db				
QY	799	AGCTCCAAGTGTGGAAGAAAGTGTAGCTCCAAGTGTGGAAGAAATTTAGCTCCAAGTGT	858	
Db				
QY	435	CCAAGGTGGACAGCTGCTCTTAATAATACATATATAGCTTTTAGGTAAATTAAGTCTAG	494	
Db				
QY	859	TGAAGAAAGTGTAGCTCCAACTGTTGAAGAAATTTAGTTCCTCAAGTGTGGAAGAAAGTGT	918	
Db				
QY	495	TGAGCGGAATCCCGGTTTGACTATTGGTTCAGATCCAGATGTATTAGGACCAATAACGCT	554	
Db				
QY	919	AGCTCCAAGTGTGGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGT	978	
Db				
QY	555	TGCAGGAATATAGATGGAGGAGGTATATACTGCATACATACAGATGCTGCCATTAACGG	614	
Db				
QY	979	TGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGT	1038	
Db				
QY	615	AACAATAGGTAATTAATCCGCGACCTCAATTAAGCATTTGGAGCAAGTACGGTTTCTCT	674	
Db				
QY	1039	AGCTGAAATGTTGAAGAAATCGTAGCTCCAAGTGTGGAAGAAATCGTAGCTCCAAGTGT	1098	
Db				
QY	675	TGAGGGGCGAGTTATTAAAGCCCACTACGACTAAGTTAACAA	716	
Db				
QY	1099	TGAAGAAAGTGTGCTGAAACCGTTGCCAACAAATTTATCAGA	1140	
Db				

RESULT 14
US-08-483-857-3

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 19:38:47; Search time 396 Seconds
(without alignments)

4870.603 Million cell updates/sec

Title: US-09-800-065-1

Perfect score: 1920

Sequence: 1 atggcgaatttctctaaa.....cgttgagctaaacaataag 1920

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1920	100.0	1920	10	US-09-800-065-1
2	1920	100.0	5513	10	US-09-800-065-3
3	59.2	3.1	510	10	US-09-864-761-18737
4	57.6	3.0	3666	10	US-09-137-531-13
5	57.6	3.0	3666	10	US-09-137-531-14
6	57.6	3.0	4197	10	US-09-137-531-7
7	57.6	3.0	4197	10	US-09-137-531-8
8	50.4	2.6	2614	9	US-09-822-846-491
9	50.4	2.6	6604	10	US-09-880-107-1748
10	49.4	2.6	7104	10	US-09-815-242-4580
11	49.4	2.6	7107	10	US-09-815-242-8291
12	49.4	2.6	7434	10	US-09-815-242-4761
13	49.4	2.6	7437	10	US-09-815-242-8869
14	49.4	2.6	8155	7	US-08-781-986A-63
15	47.4	2.5	457	10	US-09-864-761-3110
16	46.8	2.4	5361	9	US-09-742-096-2
17	46.8	2.4	6152	9	US-09-742-096-1
18	46.6	2.4	172637	10	US-09-805-458A-3
19	45.6	2.4	1920	10	US-09-800-065-1

c 20	45.6	2.4	5513	10	US-09-800-065-3	Sequence 3, Appli
c 21	44	2.3	1635	10	US-09-864-761-20241	Sequence 20241, A
c 22	44	2.3	1973	10	US-09-864-761-3471	Sequence 3471, Ap
c 23	43.6	2.3	1891	9	US-09-742-096-4	Sequence 4, Appli
c 24	42.8	2.2	2694	10	US-09-815-242-8671	Sequence 8671, Ap
c 25	42.8	2.2	3343	10	US-09-815-242-4781	Sequence 4781, Ap
c 26	42.6	2.2	486	10	US-09-864-761-10113	Sequence 10113, A
c 27	42.4	2.2	512	10	US-09-864-761-25347	Sequence 25347, A
c 28	42.4	2.2	575	10	US-09-864-761-8628	Sequence 8628, Ap
c 29	41.8	2.2	1075	10	US-09-864-761-19241	Sequence 19241, A
c 30	41.8	2.2	1403	10	US-09-864-761-2513	Sequence 2513, Ap
c 31	41.6	2.2	3996	9	US-10-087-464-42	Sequence 42, Appli
c 32	41.2	2.1	439	10	US-09-864-761-20174	Sequence 20174, A
c 33	41.2	2.1	520	9	US-10-184-644-332	Sequence 332, App
c 34	41.2	2.1	520	9	US-10-184-634-332	Sequence 332, App
c 35	40.8	2.1	465	10	US-09-864-761-230	Sequence 230, App
c 36	40.8	2.1	31517	10	US-09-070-927A-180	Sequence 180, App
c 37	40.4	2.1	18692	9	US-10-092-154-1682	Sequence 1682, Ap
c 38	40.4	2.1	18692	10	US-09-764-847-1682	Sequence 1682, Ap
c 39	40.2	2.1	446	10	US-09-864-761-20699	Sequence 20699, A
c 40	40	2.1	2765	9	US-09-117-447-5	Sequence 5, Appli
c 41	40	2.1	4358	7	US-08-781-986A-454	Sequence 454, App
c 42	39.8	2.1	456	10	US-09-864-761-4249	Sequence 4249, Ap
c 43	39.8	2.1	593	9	US-09-783-252-20	Sequence 20, Appli
c 44	39.6	2.1	1141	9	US-10-222-162-49	Sequence 49, Appli
c 45	39.4	2.1	2050	9	US-10-086-510-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-800-065-1

; Sequence 1, Application US/098000065

; Patent No. US20020094552A1

; GENERAL INFORMATION:

; APPLICANT: Bouyer, Donald H

; APPLICANT: Croquet-Valdes, Patricia

; APPLICANT: Stenos, John

; APPLICANT: Walker, David H

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR APPLICATION NUMBER: US 60/187,323

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1920

; TYPE: DNA

; ORGANISM: Rickettsia felis

; US-09-800-065-1

Query Match

Best Local Similarity 100.0%; Score 1920; DB 10; Length 1920;

Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGAATATTCTCTAAATATTTCAAAAGCAATTCAAAAGGCTTTAAACTGCT 60

Db 1 ATGGCGAATATTCTCTAAATATTTCAAAAGCAATTCAAAAGGCTTTAAACTGCT 60

QY 61 TTATTCCACCCTCAACCCGAGGATATGCTAAACGGTAGTGGAGTCTTGGTGTGCA 120

Db 61 TTATTCCACCCTCAACCCGAGGATATGCTAAACGGTAGTGGAGTCTTGGTGTGCA 120

QY 121 AGAACCGTAACCTGCTGATGGTGCAGAGCTTCGAGCGGGAACAATATAGTCTCTGGAGCC 180

Db 121 AGAACCGTAACCTGCTGATGGTGCAGAGCTTCGAGCGGGAACAATATAGTCTCTGGAGCC 180

QY 181 GTGCTTTTGTAGCGGGTTCTACTTACAAATATACCGGTGCTTTTACGGTTACTGATGCT 240

Db 181 GTGCTTTTGTAGCGGGTTCTACTTACAAATATACCGGTGCTTTTACGGTTACTGATGCT 240

Applicant

```
QY 241 GACGTAAGTGTGCTGCAATTAGATTTAAATTAATTTGACGACAGGTCTTTTTTCAGTAAC 300
Db 241 GACGTAAGTGTGCTGCAATTAGATTTAAATTAATTTGACGACAGGTCTTTTTTCAGTAAC 300
QY 301 GGTGATATTTTCATTAAGTGTGCTGATACGAGGAGGCTAATAAATTCAGTGAATTAAT 360
Db 301 GGTGATATTTTCATTAAGTGTGCTGATACGAGGAGGCTAATAAATTCAGTGAATTAAT 360
QY 361 ATTGATGATGTTTAACTTTAACTTTAACTGAGTACCGGTACTGACGCTACGGTGCAAA 420
Db 361 ATTGATGATGTTTAACTTTAACTTTAACTGAGTACCGGTACTGACGCTACGGTGCAAA 420
QY 421 CCTCGGTTGTTTCCAAAGTGGACAGTGTCTGCTTAATAATACATATACCTGCTTTAGGT 480
Db 421 CCTCGGTTGTTTCCAAAGTGGACAGTGTCTGCTTAATAATACATATACCTGCTTTAGGT 480
QY 481 AATAAATCTCTAGTGTGAGGCAATGCCGGTTTGACTATGCTTCAGATCCAGATGTATTA 540
Db 481 AATAAATCTCTAGTGTGAGGCAATGCCGGTTTGACTATGCTTCAGATCCAGATGTATTA 540
QY 541 GGACCAATAACGCTTGCAGGAATATAGATGGAGGAGGTATATAAATGACAAATACAGAT 600
Db 541 GGACCAATAACGCTTGCAGGAATATAGATGGAGGAGGTATATAAATGACAAATACAGAT 600
QY 601 GCTGCCATTAACGGAACAATAGTGAATTAATCCCGCAGCTCAAAATAGCAATTCGAGCA 660
Db 601 GCTGCCATTAACGGAACAATAGTGAATTAATCCCGCAGCTCAAAATAGCAATTCGAGCA 660
QY 661 AGTACGCTTCTCTGTGAGGCGGAGTTATTAAGCCACTACGACTAAGTTAAACAATGCA 720
Db 661 AGTACGCTTCTCTGTGAGGCGGAGTTATTAAGCCACTACGACTAAGTTAAACAATGCA 720
QY 721 GCGCCGTTATTAACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
Db 721 GCGCCGTTATTAACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
QY 781 ACAGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 781 ACAGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 AATATAGTGAATACAAATTCATTAAGCCACTAAGTGTAGGACAGGTACGGTACGTTA 900
Db 841 AATATAGTGAATACAAATTCATTAAGCCACTAAGTGTAGGACAGGTACGGTACGTTA 900
QY 901 GGGGAGCGGTTATTAAGCCACTACGACTAAGTTAAACAATGCAATGCAATGCAATGCA 960
Db 901 GGGGAGCGGTTATTAAGCCACTACGACTAAGTTAAACAATGCAATGCAATGCAATGCA 960
QY 961 CTTACAATGCAGTATTAACAGGTGCTGTTGATTAACACACAGCGGTGATATGAGGT 1020
Db 961 CTTACAATGCAGTATTAACAGGTGCTGTTGATTAACACACAGCGGTGATATGAGGT 1020
QY 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGGAATATAGTAAATCAAAATTC 1080
Db 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGGAATATAGTAAATCAAAATTC 1080
QY 1081 TTAGCCACGATAAATATAGGAGCAGGTGTGGCTTACCTTGGATGGAGCGGTATTAAGCT 1140
Db 1081 TTAGCCACGATAAATATAGGAGCAGGTGTGGCTTACCTTGGATGGAGCGGTATTAAGCT 1140
QY 1141 ACTACGACTAAGTTAAACAGATGATGCTCAGTATTAATTTTACGAATCCTGTAGTAGTA 1200
Db 1141 ACTACGACTAAGTTAAACAGATGATGCTCAGTATTAATTTTACGAATCCTGTAGTAGTA 1200
QY 1201 ACCGGAGCAATAGATAATACCGGTAATGCGCAATAAAGGTGTGTAATCTTTACCGGAGCA 1260
Db 1201 ACCGGAGCAATAGATAATACCGGTAATGCGCAATAAAGGTGTGTAATCTTTACCGGAGCA 1260
QY 1261 AGTACGGTAACCGATAATATAGGTAAACGGCAGTATTAAGCAGAGGTAAAGCGTAGGAGCA 1320
Db 1261 AGTACGGTAACCGATAATATAGGTAAACGGCAGTATTAAGCAGAGGTAAAGCGTAGGAGCA 1320
```

```
QY 1321 GGTTCGTCGAATACAAAGCGGAGTAGTAAAGCGAATGCAATAAACTTTAACGGATAAT 1380
Db 1321 GGTTCGTCGAATACAAAGCGGAGTAGTAAAGCGAATGCAATAAACTTTAACGGATAAT 1380
QY 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGCTAACAGGTAGTAGGTGGTACAGAA 1440
Db 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGCTAACAGGTAGTAGGTGGTACAGAA 1440
QY 1441 CTCCTCCACAGTGAATATAGGAGCAGGAATAACATTACGAGCCGAGGAAGCCTAGCT 1500
Db 1441 CTCCTCCACAGTGAATATAGGAGCAGGAATAACATTACGAGCCGAGGAAGCCTAGCT 1500
QY 1501 GCGAATAATATAGATTTTGGAGCTGCGAGTAATTTAGAGTTTAAACGGTCTCTGCCGTAAG 1560
Db 1501 GCGAATAATATAGATTTTGGAGCTGCGAGTAATTTAGAGTTTAAACGGTCTCTGCCGTAAG 1560
QY 1561 AATTATTAACCTTAATCGGAATATAGCAACCGGTAATTAATGCTACACCTAATATTAATGCT 1620
Db 1561 AATTATTAACCTTAATCGGAATATAGCAACCGGTAATTAATGCTACACCTAATATTAATGCT 1620
QY 1621 GCTGGTACAGTGAATGCAAAATGATGTTAGTATAGGCACAGTTGACAAATTAACATTCAA 1680
Db 1621 GCTGGTACAGTGAATGCAAAATGATGTTAGTATAGGCACAGTTGACAAATTAACATTCAA 1680
QY 1681 AATAAATAGATTTTGTAAATGCTTAAGACGCTGATGTTGATATATTAGAGCTCAG 1740
Db 1681 AATAAATAGATTTTGTAAATGCTTAAGACGCTGATGTTGATATATTAGAGCTCAG 1740
QY 1741 GCGATTAGTTTAAAGGAGCAGCTTACGCTCTTTCTTAGCTAACGTTAGCTACAGATG 1800
Db 1741 GCGATTAGTTTAAAGGAGCAGCTTACGCTCTTTCTTAGCTAACGTTAGCTACAGATG 1800
QY 1801 ATAGAGTTATCACTCTTAAAAATCATTTACCGGCTCTTGTCTAACGGTGGTGGAGTTAA 1860
Db 1801 ATAGAGTTATCACTCTTAAAAATCATTTACCGGCTCTTGTCTAACGGTGGTGGAGTTAA 1860
QY 1861 TTCTTTTGGAGTCCACAAACTTTATGACCGTACAAAGGTGACGTTGGAGCTTAAACATAG 1920
Db 1861 TTCTTTTGGAGTCCACAAACTTTATGACCGTACAAAGGTGACGTTGGAGCTTAAACATAG 1920
```

RESULT 2

US-09-800-065-3

; Sequence 3, Application US/09800065

; Patent No. US20020094552A1

; GENERAL INFORMATION:

; APPLICANT: Bouyer, Donald H

; APPLICANT: Crocquet-Valdes, Patricia

; APPLICANT: Stenos, John

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR APPLICATION NUMBER: US 60/187,323

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 5513

; TYPE: DNA

; ORGANISM: Rickettsia felis

US-09-800-065-3

Query Match 100.0%; Score 1920; DB 10; Length 5513;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGAATATTTCTCTAAATTTATTTCAAAAGCAATTTCAAAAGGTCTTTAAACTGCT 60

Db 478 ATGGCGGAATATTTCTCTAAATTTATTTCAAAAGCAATTTCAAAAGGTCTTTAAACTGCT 537

QY 61 TTATTCACCACCTCAACCCGCGGATAATGCTAACCGGTAGTGGAGTCTCTTGGTGTGCA 120


```

; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18737
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
; OTHER INFORMATION: NT HIT: U01287.1, EVALUATE 1.30e+00
US-09-864-761-18737

```

```

Query Match 3.1%; Score 59.2; DB 10; Length 510;
Best Local Similarity 46.1%; Pred. No. 0.0007;
Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 104 GAGTCCTGGTGCAGACCGTAACTGCTGATGCTGACAGCTTGCAGCGGAAACAA 163
DB 510 GTGTTGATGCTGATGCGAGGACCGATCTGAAGCAGGTGTTGATGCGAGGTGCTGAAAAAG 451

QY 164 ATATAGTCTGCGACCGGCTTTTCTAGCGGTTCTACTTTACAATATACCGTCTT 223
DB 450 GTGTTGCGAGGTCAGATGCTGCTGCTGATGCTGATGCTGAGCGCGCTGATGCTG 391

QY 224 TTACGGTTACTGATGCTGACGTAAGTCTTCGTGATTTAGATTTAAATTAATTTTGCAGCAG 283
DB 390 ATGCAGGTATTTGTTGTTGTCAGATGTTGATGTCAGATGTTGAAGCAGGTGCTGAAAAAG 331

QY 284 GTCTTTTTCAGTAACCTGGTATTTTCATTTAGTTTCAGTTCAGTTCAGCGGAGGAGCTA 343
DB 330 GTGTTGATGCGAGGTCGATGCTGCTGATGCTGATGCTGATGCTGATGCTGCTGCTG 271

QY 344 ATAACTTCAGTTAATTAATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
DB 270 CAGATGTTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 211

QY 404 CAGCCTACGCTGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 463
DB 210 CAAGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 151

```

```

QY 464 CATATACCTGCTTTAGTAAATATACTCTAGGTGGAGCAATGCCGGTTTGACTATTGCTT 523
DB 150 TTGATGCAGATGTTGAAGCGGGTGTGTGATGCTGATGCTGAGGCTGCTGAAGCAGGTATTG 91
QY 524 CAGATCCAGATG 535
DB 90 CTGATACAGATG 79

RESULT 4
US-09-137-531-13
; Sequence 13, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EP0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/692,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-137-531-13

```

```

Query Match 3.0%; Score 57.6; DB 10; Length 3666;
Best Local Similarity 47.0%; Pred. No. 0.0042;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

QY 116 CTGCAAGAACCGTAACCTGCTGATGGTGCAGAGCTTGCAGCGGAAACAAATATAGTCTG 175
DB 2735 CTTTAGCAATTTTCATTAGCTGATGACATCTTAATGTAAGTGCACAACTGTTGATACTG 2794

QY 176 GAGCGGTGCTTTTGTAGCGGGTCTTACTTTACAATATACCGGTGCTTTTACGGTTACTG 235
DB 2795 CAACTGTTTCATTAAAG-----ATAGTGCATAATTAATTCATTATCTCTTACATTAGTTG 2848

QY 236 ATGCTGACGTAAGTGTTCGTCATTAGATTTAAATTAATTTTGCAGCAGGTCTTTTTCAG 295
DB 2849 AACTGCTGCTAATACAGGTGTATTTGCTACAACCTGTTCAAGCTGGTACATTATCTTCT 2908

QY 296 TAATCGTGTATTTTCATTAGTTTCAGTGTAGTAGTGCAGGAGGAGCTAATAAATCTTCAG 355
DB 2909 TAACTGCTGTACATTAACAGTTTACCTTATGTCAGATGCTGCAAGTGTGCTGCTGCTG 2968

QY 356 TTAATATTGATGCTTTTAACTTAACTTTACAGGTACCGGTACTGACGCTACCGGTG 415
DB 2969 AAAATATTACTGCTAGCTAAACATTAAAGAAAC---TACTGGAGCAATTTACTTCTGATA 3025

QY 416 CAAATCCTGCTGTTTATTCCAAAGTGGACAGCTGCTGCTAATAATATACATATACCTT 475
DB 3026 CATTTACACAAGGTGATTACCATCAGCAGCTACACAGCTGAATATCTTCTTAATCAA 3085

QY 476 TAGGTAATACTAGTGGAGCGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG 535
DB 3086 TTGCTCGATTATACATTTGCAACAGGTGAAGTACCTTTTAAATATTTGATAATGCTG 3145

QY 536 TATTAGGACCAATAACGCTTGCAGGAAATATAGATGAGGAGGTATTAATAACTGACAATA 595

```

Db 3146 GTGCTCAAGTAATTAACCTTAGCAGTAAAGGTCACAAAGGT--GTAGCTGATGCTA 3202
QY 596 CAGATGCTGCCATTAACGGAACAATAGTAATACTATCCGGCAGCTCAATAAATGACATTG 655
Db 3203 TCAATGCTACATTTGACAGGTACTGCAACTGTTCTGGAGACAAAGTAGTTATTAATACAG 3262
QY 656 GAGCAAGTACGCTTCTCTGGAGGGCAGTTATTAAGCCACATACGACTAAGTTAAACAA 715
Db 3263 CTACAAACAGGTGTTGGTCTGAAGTTGAAGTTACATTTCTCTCTGTTAATCAAGTATTAA 3322
QY 716 ATGCAG 721
Db 3323 ATGCAG 3328
RESULT 5
US-09-137-531-14
; Sequence 14, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-09-137-531-14
Query Match 3.0%; Score 57.6; DB 10; Length 3666;
Best Local Similarity 47.0%; Pred. No. 0.0042;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;
QY 116 CTGCAAGACCGCTAACTGCTGATGGTGACAGCTTGCAGCCGCGAACAATATAGTCTCTG 175
Db 2735 CTTTAGGAATTCATTAGCTGATGACAGATCTTAATGTAGTGAACAACTGTTGATACGTG 2794
QY 176 GAGCCGGTCTTTGTAGCGGGTCTTACATTAACAATATACCGGTGCTTTTACGGTTACTG 235
Db 2795 CAACTGTTTCATTAAG-----ATAGTGCAAAATTAATTCATTCTTACATTTAGTTG 2848
QY 236 ATGCTGACGTAAGTGTCTGCTGATTAAGTAAATTAATTTTGCAGCAGCTCTTTTTCAG 295
Db 2849 AACTGGTCTAATACAGGTGATTGTTACAACCTGTTCAAGCTGTTACATATCTTCTT 2908
QY 296 TAACTGGTATATTCATTAGTTGCTGATAGTATACGGGAGGAGGTAATAACTGTCAG 355
Db 2909 TAACTGGTATATTCATTAGTTGCTGATAGTATACGGGAGGAGGTAATAACTGTCAG 355
QY 356 TAAATATGATGATGTTTAACTTAACTTTAAACAGGTACCGGTACTGCGAGCTACGGTG 415
Db 2969 AATATATTACTGCTACGGTACATTAAGAAAC---TACTGGCAATTAATCTCTGATA 3025
QY 416 CAAATCCTGCGTGTATTTCRAAGGTGGAAGCTGCTGTAATAATACATATACATGCTT 475

Db 3026 CATTTACAAAGGTGTATTACCATCAGCAGCTACAGCAGCTGAATATATCTTAATCAA 3085
QY 476 TAGTAATAATACTCTAGTGGAGGGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG 535
Db 3086 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACCTTTAAATATTGATATGCTG 3145
QY 536 TATTAGGACCAATAACGCTTGCAGGAATAATAGATGGAGGAGGTATAATAACTGCACAATA 595
Db 3146 GTGCTCAAGTAATTAACCTTAGCAGTAAAGGTCACAAAGGT--GTAGCTGATGCTA 3202
QY 596 CAGATGCTGCCATTAACGGAACAATAGTAATACTATCCGGCAGCTCAATAAATGACATTG 655
Db 3203 TCAATGCTACATTTGACAGGTACTGCAACTGTTCTGGAGACAAAGTAGTTATTAATCAG 3262
QY 656 GAGCAAGTACGCTTCTCTGGAGGGCAGTTATTAAGCCACATACGACTAAGTTAAACAA 715
Db 3263 CTACAAACAGGTGTTGGTCTGAAGTTGAAGTTACATTTCTCTCTGTTAATCAAGTATTAA 3322
QY 716 ATGCAG 721
Db 3323 ATGCAG 3328
RESULT 6
US-09-137-531-7
; Sequence 7, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-137-531-7
Query Match 3.0%; Score 57.6; DB 10; Length 4197;
Best Local Similarity 47.0%; Pred. No. 0.0044;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;
QY 116 CTGCAAGACCGCTAACTGCTGATGGTGACAGCTTGCAGCCGCGAACAATATAGTCTCTG 175
Db 2919 CTTTAGGAATTCATTAGCTGATGACAGATCTTAATGTAGTGAACAACTGTTGATACGTG 2978
QY 176 GAGCCGGTCTTTGTAGCGGGTCTTACATTAACAATATACCGGTGCTTTTACGGTTACTG 235
Db 2979 CAACTGTTTCATTAAG-----ATAGTGCAAAATTAATTCATTCTTACATTTAGTTG 3032
QY 236 ATGCTGACGTAAGTGTCTGCTGATTAAGTAAATTAATTTTGCAGCAGCTCTTTTTCAG 295
Db 3033 AACTGGTCTAATACAGGTGATTGTTACAACCTGTTCAAGCTGTTACATATCTTCTT 3092
QY 296 TAACTGGTATATTCATTAGTTGCTGATAGTATACGGGAGGAGGTAATAACTGTCAG 355
Db 3093 TAACTGGTATATTCATTAGTTGCTGATAGTATACGGGAGGAGGTAATAACTGTCAG 355
QY 356 TTAATATTGATGATGTTTAACTTAACTTTAAACAGGTACCGGTACTGCGAGCTACGGTG 415

Db 3153 AAAAATATTACTGCTAGCGTAACATTAAAGAAAC---TACTGGAGCAATTAATCTCTGATA 3209
 Qy 416 CAAATCTCGTGTGTTATTCCAAAGTGGACAGCTGCTGCTAAATATACATATATCTGCT 475
 Db 3210 CATTACACAGGTTATTAACCATCAGCAGCTACAGCAGCTGAATATATCTTCTAAATCAA 3269
 Qy 476 TAGGTAATATACTCTAGGTGGAGGCAATGCGGTTTGACTATATGCTTCAGATCCAGATG 535
 Db 3270 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACITTAATATGTAATGCTG 3329
 Qy 536 TATTAGGACCATACGCTTGCAGGAATATAGATGGAGGAGGTATATACTGACAATA 595
 Db 3330 GTGCTCAAGTAAATTAACCTAGCAGGTAAAGGAGGTGCAAGGT---GTAGCTGATGCTA 3386
 Qy 596 CAGATGCTGCCATTAACGGAACAATAGGTAACTAATCTCCGCGACGCTCAAAATACGATTG 655
 Db 3387 TCAATGCTACATTTCCAGGTACTGCACTGTTCTGGAGACAAGTAGTATTAATCAG 3446
 Qy 656 GAGCAAGTACGCTTCTCTGGAGGGGAGTATTAAAGCCACTACGACTAAGTTAAACAA 715
 Db 3447 CTACACAGAGGTGTTGGTTCTGAAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3506
 Qy 716 ATGCAG 721
 Db 3507 ATGCAG 3512

RESULT 7

US-09-137-531-8
 ; Sequence 8, Application US/09137531
 ; Patent No. US20020048816A1
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/137,531
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/682,517
 ; FILING DATE:
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4197 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Bacillus sphaericus
 ; INDIVIDUAL ISOLATE: P-1
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 95..3850
 ; FEATURE:
 ; NAME/KEY: mat_peptide
 ; LOCATION: 185..3850
 ; FEATURE:
 ; NAME/KEY: sig_peptide
 ; LOCATION: 95..184
 ; US-09-137-531-8

Query Match 3.0%; Score 57.6; DB 10; Length 4197;
 Best Local Similarity 47.0%; Pred. No. 0.0044;
 Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

Qy 116 CTGAAGAACCCTAACTGCTGATGTCAGAGCTTGCAGCGGACCAACAAATATAGTCTCTG 175
 Db 2919 CTTTAGGAATTTCAATGCTGATGCAGATCTTAATGTAAAGTGAAGTGCACAACTGTTGATACGT 2978
 Qy 176 GAGCCGCTGCTTTTGTAGCGGTTCTACTTTTCAATATACCGGTGCTTTTACGGTTACTG 235
 Db 2979 CAACCTGTTTCAATAAAG-----ATAGTGAATAATTCATTATCTTACATTAAGTTG 3032
 Qy 236 ATGCTGACCTAAGTGTGCTGCATTAATTAATAATTTTGCAGCAGGTCTTTTTCAG 295
 Db 3033 AAATCTGCTGCTAATACAGGTGTTTGTACAACCTGTTTCAAGCTGCTACATTAATCTTCTT 3092
 Qy 296 TAACTGCTGATATTTCATTAGTTTCAGTGTAGTACGAGGAGGAGCTAAATAACTTCGAG 355
 Db 3093 TAACTGCTGCTACATTAACAGTTTACTTATGCAGATGCTAAATAATGCTGCAGGTGTTGCTG 3152
 Qy 356 TTAATATTGATGTTGTTTAACTTAACTTTTACAGAGGTACCGGTACTGACGCTACGCTG 415
 Db 3153 AAAATATTACTGCTAGCGTAACATTAAGAAAC---TACTGGAGCAATTAATCTTCTGATA 3209
 Qy 416 CAAATCTCGCTGTTGTTTCAAGGTGGACAGGTGCTGCTAAATAATACATATACATGCTT 475
 Db 3210 CATTTACACAGGTGTTATACCATCAGCAGCTACAGCAGCTGAATATACTTCTAAATCAA 3269
 Qy 476 TAGTAATATAACTCTAGTGGAGCGAATCCCGTTTTCAGTATTGCTTCAGATCCAGATG 535
 Db 3270 TTGCTGCAGATTATACATTTTGAACAGGTGAAGGATTCACITTAATATATGATAATGCTG 3329
 Qy 536 TATTAGGACCAATAACGCTTGCAGGAATATAGATGGAGGAGGTATTAATACATGACAATA 595
 Db 3330 GTGCTCAAGTAAATTAACCTAGCAGGTAAAAAGGTGCAAGGT---GTAGCTGATGCTA 3386
 Qy 596 CAGATGCTGCCATTAACGGAACAATAGTAACTAATCTCCGCGACGCTCAATAGCATG 655
 Db 3387 TCATGCTACATTTGCGAGGTACTGCAACTGTTCTCGAGACAAGTAGTATTAAATCAG 3446
 Qy 656 GAGCAAGTACGCTTTCTCTGGAGGGGCGAGTTATTAAAGCCACTACGACTAAGTTAAACAA 715
 Db 3447 CTACACAGAGGTGTTGGTTCTGAAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3506
 Qy 716 ATGCAG 721
 Db 3507 ATGCAG 3512

RESULT 8

US-09-822-846-491/c
 ; Sequence 491, Application US/09822846
 ; Publication No. US20030027139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: Lavallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Steininger II, Robert J.
 ; APPLICANT: Bowman, Michael R.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fechtel, Kim
 ; APPLICANT: Howes, Steven H.
 ; APPLICANT: Resnick, Richard J.
 ; APPLICANT: Gulukota, Kamalakar
 ; APPLICANT: Graham, James R.
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 ; FILE REFERENCE: GIN 6400
 ; CURRENT APPLICATION NUMBER: US/09/822,846
 ; CURRENT FILING DATE: 2001-03-29

```

; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 491
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-491

Query Match      2.6%; Score 50.4; DB 9; Length 2614;
Best Local Similarity 47.7%; Pred. No. 0.15;
Matches 147; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 125 CGTAACTGCTGATGCTGACAGCTTGCAGCGGAGCAAAATAGTCTCGAGCCGGT 184
Db 2295 CTGTTGCTGCTGCTGGGTCTGTAAGCCCTGGCGCTGGAACCTGGGCTGGAGTGGG 2236

QY 185 CTTTGTAGCGGGTCTACTTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACG 244
Db 2235 CTGGGGTTGGGAGGAGCGGCTGTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 2176

QY 245 TAAAGTGTGCTGATAGATTAATAATTTTGCAGCAGGCTCTTTTTCAGTAACTGGTG 304
Db 2175 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2116

QY 305 ATATTTCATTAGTTCAGTGTAGATACGGGAGGAGCTAAATAAACTTGCAGTTAATATG 364
Db 2115 CTGCGCAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2056

QY 365 ATGATGGTTTAACTTAACTTTTAAACAGGTACCGGTACTGCAGCCTACGGTGCAAAATCTG 424
Db 2055 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1996

QY 425 CGTTGTTA 432
Db 1995 CTCAGGTA 1988

RESULT 9
US-09-880-107-1748/c
; Sequence 1748, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1748
; LENGTH: 6604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D83783
US-09-880-107-1748

Query Match      2.6%; Score 50.4; DB 10; Length 6604;
Best Local Similarity 47.7%; Pred. No. 0.23;
Matches 147; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 125 CGTAACTGCTGATGCTGACAGCTTGCAGCGGAGCAAAATAGTCTCTCGAGCCGGT 184
Db 6291 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6232

; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 491
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-491

Query Match      2.6%; Score 49.4; DB 10; Length 7104;
Best Local Similarity 44.2%; Pred. No. 0.4;
Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAACACTTACAAATGCAATGCAAGTATTAACAGGTGCTGTGATAACACCACAG 784
Db 3809 CTGTGCAACTCAAAATCAAGCAATGATAATACAACCTGGTGTCTACAACTGAAGAAAAA 3868

QY 785 GCGGTGATGATGATGCTGCTTAAATTTAAACGAGCGGTGAGTCAAGTAACTGGAATA 844
Db 3869 ATGCAGCAAAAGATTTTAGTTTTAAAGCTAAAGAAAAAGCGTATCAAGATATCTTAATG 3928

```

QY 845 TAGGTAATACAAATTCATTAGGACCAATAAGTGTAGGAGCAGGTACGGGTACGTTAGGGG 904
Db 3929 CACAAACAACACTAACGATGTACGCAAAATTAAGATCAAGCAGTTGCTGATGTTCAAGGTA 3988
QY 905 GAGCGGTTATTAAAGCCACTACGACTTAAGTTAACGAATG---CAGCGTCGGTATTAAACAC 961
Db 3989 TTACTGCAGATACACAAATTAAGATGTTGCGAAAGATGAATTACACAAAGCAAGGG 4048
QY 962 TTACAAATGACATTAACAGGTGCTGTTGATAACACACAGCGCGGTGATATGTTAGGTG 1021
Db 4049 AACAAAGAACCTTATTCACAACTGCAGATGCGACTACTGAAGAAAAGAACAGCAA 4108
QY 1022 TCGTAATTAAGTGGACCAATGCTCAAGTAACCGGGAATATAGGTATACAAATTCAT 1081
Db 4109 ATCAACAGTAGATGCAAAATTAAGCAAGGTAATCAAAATATTGAAATGTCACAGTCAA 4168
QY 1082 TAGCCACGATAAATATAGGAGCAGGTGCTGCTACCTTGGATGAGCGGTATTAAAGGTA 1141
Db 4169 TCGATGATGTAACACTGCAAAAGATAATGCAATTCAGCAATTCACCAATTCAGCAT 4228
QY 1142 CTAGCACTAAGTTACAGATGATGCGTCAAGTATTATTAATTTACGAATCCTGTAGTAGTAA 1201
Db 4229 CAACAGATGTTAAAACGAATGCAAGAGCGGAATTCCTAACTCAATGCAAAATAAAATAA 4288
QY 1202 CCGGAGCAATAGATAATACCGGTAATGCCAATAAGGTGTGGTATCTTTTACCGGAGCAA 1261
Db 4289 CTGAATCTTAATTAATGAGACTACTAATGAAGAAAAGGTAAACGATATTGGACCA 4348
QY 1262 GTACCGTTAACCGATAATATAGGT 1284
Db 4349 TTAGAGCTGCATATGAAGAAGGT 4371

RESULT 11

US-09-815-242-8291
; Sequence 8291, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8291

LENGTH: 7107

TYPE: DNA

ORGANISM: Staphylococcus aureus

FEATURE:

; NAME/KEY: CDS
; LOCATION: (1)...(7107)
US-09-815-242-8291

Query Match

Best Local Similarity 2.6%; Score 49.4; DB 10; Length 7107;
Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAAACACTTACAAATGCAATGCAATTAACAGGTGCTGTTGATACACACACAC 784
Db 3809 CTGTGACAACTCAAAATCAAGCAATTTGATAATACAACTGGTCTACAACTCAAGAAAAA 3868
QY 785 CGGTGATGATGATGCTCTTAATTTAAACGGAGCGCTTGAGTCAACTAACTGAGAAATA 844
Db 3869 ATGACGAAAAAGATTAGTTTTTAAAGCTAAAGAAAAAGCGTATCAAGATATCTTAATG 3928
QY 845 TAGGTAATACAAATTCATTAGGCACAATTAAGTGTAGGAGCAGGTACGGCTAGTTAGGG 904
Db 3929 CACAAACAACTAAGCATGTTACGCAATTTAAAGATCAAGCAGTTGCTGATGTTCAAGTA 3988
QY 905 GAGCGGTTATTAAAGCCACTACGACTAAAGTTAAACGAATG---CAGCGTCGGTATTAAACAC 961
Db 3989 TTACTGCAGATACACAAATTAAGATGTTGCGAAAGATAATGCAATTCAGCAATTCACCAATTCAGCAT 4228
QY 962 TTACAAATGCAATTAACAGGTGCTGTTGATAACACACAGCGGTGATATGTTAGGTG 1021
Db 4049 AACAAAGACCTTATTGCACAAACTGCAGATGCGACTACTGAAGAAAAGAACAGCAA 4108
QY 1022 TCGTAATTAAGTGGAGCATTGATCAAGTAACCGGGAATTAAGGTATCAAAATATTGAAATTCACAGTCAA 1081
Db 4109 ATCAACAGTAGATGCAAAATTAAGCAAGGTAATCAAAATATTGAAATTCACAGTCAA 4168
QY 1082 TAGCCACGATAAATATAGGAGCAGGTGCTGCTACCTTGGATGAGCGGTATTAAAGGTA 1141
Db 4169 TCGATGATGTAACACTGCAAAAGATAATGCAATTCAGCAATTCAGCAATTCAGCAT 4228
QY 1142 CTAGCACTAAGTTACAGATGATGCGTCAAGTATTATTAATTTACGAATCCTGTAGTAGTAA 1201
Db 4229 CAACAGATGTTAAAACGAATGCAAGAGCGGAATTCCTAACTCAATGCAAAATAAAATAA 4288
QY 1202 CCGGAGCAATAGATAATACCGGTAATGCCAATAAGGTGTGGTATCTTTTACCGGAGCAA 1261
Db 4289 CTGAATCTTAATTAATGAGACTACTAATGAAGAAAAGGTAAACGATATTGGACCA 4348
QY 1262 GTACCGTTAACCGATAATATAGGT 1284
Db 4349 TTAGAGCTGCATATGAAGAAGGT 4371

RESULT 12

US-09-815-242-4761
; Sequence 4761, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4761
; LENGTH: 7434
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4761

```

```

Query Match      2.6%; Score 49.4; DB 10; Length 7434;
Best Local Similarity 44.2%; Pred. No. 0.41;
Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAACACTTACAAATGCAAAATGCAAGTATTAAACAGCGTGTGATTAACACACAG 784
DB 4160 CTGTGCACTCAAAATCAAGCAATTTGATAATACAACTGGTCTACAACTGAAGAGAAA 4219

QY 785 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 844
DB 4220 ATGCAGCAAAAGATTTAGTTTTAAAGCTAAAGAAAACGGTATCAAGATATCTTAATG 4279

QY 845 TAGGTAATACAATTCATTAGGCACATTAAGTAAACGAGTGTAGGAGCGGTAGGAGCG 904
DB 4280 CACAAACAACACTAATGATGTACGCAATTAAGATCAACAGCTTGTGATATTCAGGTA 4339

QY 1022 TCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGAATATAGTAATACAAATTCAT 1081
DB 4460 ATCAACAAGTAGACGCACAAATTAAACACAGGTAATCAAAATATTGAATATGCACAGTCAA 4519

QY 1082 TAGCCACGATTAATATAGGAGCGGTGTGCTACCTTGGATGGAGCGGTTATTAAAGCTA 1141
DB 4520 TCGATGATGTAACACTGCAAAAGATAATGCAATTCGAATTCACCAATTCAGACAT 4579

QY 1202 CCGGAGCAATAGATATACCGGTAAATGCCAATAAAGGTGTGTAATCTTTACCGGAGCAA 1261
DB 4640 CTGAATCTTAAATATAGACTACTAATGAAGAAAAGGTACGATATTTGGACCAG 4699

QY 1262 GTACGGTAACCGATATATAGGT 1284
DB 4700 TTAGACGACATATGAAGAAGT 4722

```

```

RESULT 13
US-09-815-242-8869
; Sequence 8869, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```

```

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8869
; LENGTH: 7437
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE: CDS
; NAME/KEY: CDS
; LOCATION: (1)...(7437)
US-09-815-242-8869

```

```

Query Match      2.6%; Score 49.4; DB 10; Length 7437;
Best Local Similarity 44.2%; Pred. No. 0.41;
Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAACACTTACAAATGCAAAATGCAAGTATTAAACAGCGTGTGATTAACACACAG 784
DB 4160 CTGTGCACTCAAAATCAAGCAATTTGATAATACAACTGGTCTACAACTGAAGAGAAA 4219

QY 785 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 844
DB 4220 ATGCAGCAAAAGATTTAGTTTTAAAGCTAAAGAAAACGGTATCAAGATATCTTAATG 4279

QY 845 TAGGTAATACAATTCATTAGGCACATTAAGTAAACGAGTGTAGGAGCGGTAGGAGCG 904
DB 4280 CACAAACAACACTAATGATGTACGCAATTAAGATCAACAGCTTGTGATATTCAGGTA 4339

QY 905 GAGCGGTTATTAAAGCCACTACGACTAAGTAAACGAATG---CAGCGTCGGTATTAAAC 961
DB 4340 TTACTGCAGATACAAATTAAGATGTTGCGAAAGATGAATAGCAACAAAGCAACG 4399

QY 962 TTACAAATCCAGTATTAAACAGGTGCTGTTGATAACACACAGCGGTGATATGTTAGGTG 1021
DB 4400 ACACAAACCGCTTATTGCAACAACTGCGAGTGGGACTACTGAAGAAAAGCAACAGCAA 4459

QY 1022 TCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGAATATAGTAATACAAATTCAT 1081
DB 4460 ATCAACAAGTAGACGCACAAATTAAACACAGGTAATCAAAATATTGAATATGCACAGTCAA 4519

QY 1082 TAGCCACGATTAATATAGGAGCGGTGTGCTACCTTGGATGGAGCGGTTATTAAAGCTA 1141
DB 4520 TCGATGATGTAACACTGCAAAAGATAATGCAATTCGAATTCACCAATTCAGACAT 4579

QY 1142 CTACGACTAAGTTAAACAGATGATGCTGCTAGTATTATATTTACGAATCTGTAGTAGTAA 1201
DB 4580 CAACAGATTTAAACGAATGCAAGAGCGGAATTTGCTGAATGCAAAATAAATAA 4539

QY 1202 CCGGAGCAATAGATATACCGGTAAATGCCAATAAAGGTGTGTAATCTTTACCGGAGCAA 1261
DB 4640 CTGAATCTTAAATATAGACTACTAATGAAGAAAAGGTACGATATTTGGACCAG 4699

QY 1262 GTACGGTAACCGATATATAGGT 1284

```

Db 4700 TTAGAGCAGCATATCAAGAAGGT 4722

RESULT 14

US-08-781-986A-63

Sequence 63, Application US/08781986A

Publication No. US20030054436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob

REGISTRATION NUMBER: 30,446

REFERENCE/DOCKET NUMBER: PB248PP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 8155 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-781-986A-63

Query Match 2.6%; Score 49.4; DB 7; Length 8155;

Best Local Similarity 44.2%; Pred. No. 0.43;

Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAAACACITACAAATGCAATGCAGTATTAAACAGGTGCTGTTGATTAACACAGGCTTTCAGTCAAGTAACTGGAATA 844

Db 6152 CTGTGACAACTCAAAATCAAGCAATTTGATAATCAAACTGGTGTACAACTGAAGAGAAA 6211

QY 785 GCGGTGATGATGAGGTGCTTAAATTTAAACGAGCGTTTCAAGTCAAGTAACTGGAATA 844

Db 6212 ATGCAGCAAGATTGTTAGTTTAAAGCTTAAGAAAAGCGTATCAAGATATCTTAAATG 6271

QY 845 TAGTAATACAAATTCATTAGCGCAATAAGTGTAGGAGCAGGTACGCGTACGTTAGGGG 904

Db 6272 CACAACAACATGATGTTACGCAAAATTAAGATCAACAGTGTCTGATATTCAAGTA 6331

QY 905 GAGCGTTTAAAGCCACTACGACTAAGTTAAGCAATG---CAGCGTCGGTATTACAC 961

Db 6332 TTACTGCAGATACAACAATTAAGATGTTGCGAAAGATGATTAAGCAACAAAAGCAACG 6391

QY 962 TTCAAAATGCGATTTAAACAGGTGCTGTTGATAACACACAGCGGTGATAATGTTAGTG 1021

Db 6392 AAAAAAGCGCTATTGTCACAACTCCAGTGGGACTCTGAAGAAAAAGCAACAA 6451

QY 1022 TCCTAAATCTAAGTGGAGCATTTAGTCAAGTAACCGGGAATATAGGTTAATACAAATTCAT 1081

Db 6452 ATCAACAAGTAGCGCAATTAACACAAGGTATCAAAATATTGAAAAATGCACAGTCAA 6511

QY 1082 TAGCCACGATAAATATAGGAGCAGGTGTGGCTACCTTGGATGGAGCGGTTTATTAAAGCTA 1141

Db 6512 TCGATGATGTAACACACTGCAAAAGATATATGCAATTCGAATTCACCAATTCGAAGCAT 6571

QY 1142 CTACGACTAAGTTAACAGATGATCGGCAGTATTAAATTTACGAATCTCTAGTAGTAA 1201

Db 6572 CAACAGATGTTAAACGAATGCAAGAGCGGAATTCCTAACTGAAATGCAAAATGCAAAATAA 6631

QY 1202 CCGGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGTGGTAACTCTTTACCGGAGCAA 1261

Db 6632 CTGAATACTTAATAATAATGAGACTACTAATGAAGAAAAAGTAAACGATATTGGACCAG 6691

QY 1262 GTACGGTAACCGATAATATAGGT 1284

Db 6692 TTAGAGCAGCATATGAAGAAGGT 6714

RESULT 15

US-09-864-761-3110/c

Sequence 3110, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 3110

LENGTH: 457

TYPE: DNA

ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
US-09-864-761-3110

Query Match 2.5%; Score 47.4; DB 10; Length 457;
Best Local Similarity 52.2%; Pred. No. 0.3;
Matches 105; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 103 GGAGTCCTGGTGGCTGCAAGAACCGTAACCTGATGGTGACAGCTTGCAGCCGGAACA 162
DB 103 GGAGTCCTGGTGGCTGCAAGAACCGTAACCTGATGGTGACAGCTTGCAGCCGGAACA 162
QY 201 GGTGTTGATGCTGATGCAGGTACCGATGCTGAAGCAGGTGTTGATGCAGGTGCTGAAAAA 142
DB 201 GGTGTTGATGCTGATGCAGGTACCGATGCTGAAGCAGGTGTTGATGCAGGTGCTGAAAAA 142
QY 163 AATATAGTCTCTGGAGCGGCTCTTTTGTAGCGGTTCTTACTTTACAATATACCGTGCT 222
DB 163 AATATAGTCTCTGGAGCGGCTCTTTTGTAGCGGTTCTTACTTTACAATATACCGTGCT 222
QY 141 GGTGTTGGTGCAGTGCAGATGCTGATGCTGCTAGTAAATGCTGATGCAGGCGCTGATGCT 82
DB 141 GGTGTTGGTGCAGTGCAGATGCTGATGCTGCTAGTAAATGCTGATGCAGGCGCTGATGCT 82
QY 223 TTTACGGTTACTGATGCACGTAACTGCTTCGTCATTAGATTTAAATAATTTTCAGCA 282
DB 223 TTTACGGTTACTGATGCACGTAACTGCTTCGTCATTAGATTTAAATAATTTTCAGCA 282
QY 81 GATCAGGTATTGGTGGTGGTGCAGATGTTGATGCAGATGTTGAAGCAGGTGCTGAAAAA 22
DB 81 GATCAGGTATTGGTGGTGGTGCAGATGTTGATGCAGATGTTGAAGCAGGTGCTGAAAAA 22
QY 283 GGTCTTTTTTTCAGTAACCTGGT 303
DB 283 GGTCTTTTTTTCAGTAACCTGGT 303
QY 21 GGTGTTGATGCAGGTGCAGAT 1
DB 21 GGTGTTGATGCAGGTGCAGAT 1

Search completed: April 17, 2003, 19:48:13
Job time : 558 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 13:33:35 ; Search time 8577 Seconds
(without alignments)
5628.259 Million cell updates/sec

Title: US-09-800-065-1
Perfect score: 1920
Sequence: 1 atggcgaatttctctaa.....cgttgagcgttaaacataag 1920

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 24791104 seqs, 12571243825 residues
Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1920	100.0	1920	31	US-09-800-065-1
2	1920	100.0	5513	31	US-09-800-065-3
3	597.6	31.1	7088	4	US-08-029-329-1
4	91.2	4.8	7814	66	US-60-226-176-1986
5	91.2	4.8	7814	67	US-60-233-468-1986
6	91.2	4.8	7814	75	US-60-313-371-1986
c 7	76.8	4.0	7814	66	US-60-226-176-1986
c 8	76.8	4.0	7814	67	US-60-233-468-1986
c 9	76.8	4.0	7814	75	US-60-313-371-1986
c 10	74.2	3.9	540	1	PCT-US00-41008-1
c 11	74.2	3.9	186935	66	US-60-226-176-1976
c 12	74.2	3.9	186935	67	US-60-233-468-1976
c 13	74.2	3.9	186935	75	US-60-313-371-1976
c 14	71.2	3.7	961	80	US-60-360-207-31796
15	69.6	3.6	186935	66	US-60-226-176-1976
16	69.6	3.6	186935	67	US-60-233-468-1976
17	69.6	3.6	186935	75	US-60-313-371-1976
c 18	63.8	3.3	486	1	PCT-US00-22496-13
c 19	63.8	3.3	486	24	US-09-639-207-13
c 20	60.4	3.1	7218	8	US-08-466-194-14
c 21	59.2	3.1	510	1	PCT-US01-00663-15232

c 22 59.2 3.1 510 33 US-09-864-761-18737 Sequence 18737, A
c 23 59.2 3.1 510 41 US-10-182-993-14763 Sequence 14763, A
c 24 59.2 3.1 510 41 US-10-182-995-11883 Sequence 11883, A
c 25 59.2 3.1 510 42 US-10-203-134-15069 Sequence 15069, A
c 26 59.2 3.1 510 42 US-10-203-136-15065 Sequence 15065, A
c 27 59.2 3.1 510 42 US-10-203-137-15232 Sequence 15232, A
c 28 59.2 3.1 510 42 US-10-203-138-7146 Sequence 7146, A
c 29 59.2 3.1 510 42 US-10-203-139-14660 Sequence 14660, A
c 30 59.2 3.1 510 67 US-60-236-359-9553 Sequence 9553, A
c 31 57.6 3.0 604 64 US-60-207-458-24992 Sequence 24992, A
c 32 57.6 3.0 3666 15 US-09-137-531-13 Sequence 13, Appl
c 33 57.6 3.0 3666 15 US-09-137-531-14 Sequence 14, Appl
c 34 57.6 3.0 4197 15 US-09-137-531-7 Sequence 7, Appl
c 35 57.6 3.0 4197 15 US-09-137-531-8 Sequence 8, Appl
c 36 57.4 3.0 390 21 US-09-540-234-2089 Sequence 2089, A
c 37 56.8 3.0 961 80 US-60-360-207-31796 Sequence 31796, A
c 38 56.8 3.0 1954 80 US-60-360-207-4707 Sequence 4707, A
c 39 56.8 3.0 8307 69 US-60-258-275-230 Sequence 230, App
c 40 56 2.9 632 33 US-09-865-439A-30490 Sequence 30490, A
c 41 56 2.9 632 64 US-60-207-458-74787 Sequence 74787, A
c 42 56 2.9 897 80 US-60-360-207-39592 Sequence 39592, A
c 43 55.8 2.9 1118 80 US-60-360-207-33054 Sequence 33054, A
c 44 55.4 2.9 1236 39 US-10-077-584-3 Sequence 3, Appl
c 45 55.4 2.9 1236 70 US-60-269-157-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-800-065-1

; Sequence 1, Application US/098000065

; GENERAL INFORMATION: Bouyer, Donald H

; APPLICANT: Crocquet-Valdes, Patricia

; APPLICANT: Stenos, John

; APPLICANT: Walker, David H

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1920

; TYPE: DNA

; ORGANISM: Rickettsia felis

; US-09-800-065-1

Query Match 100.0%; Score 1920; DB 31; Length 1920;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGGAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGCTCTTAAACCTGCT 60
DB 1 ATGCGGAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGCTCTTAAACCTGCT 60
QY 61 TTATTCACCACTCAACCGGAGCGATATGCTAACCGGTAGTGAGTCTTGGTGCTGCA 120
DB 61 TTATTCACCACTCAACCGGAGCGATATGCTAACCGGTAGTGAGTCTTGGTGCTGCA 120
QY 121 AGAACCTTAAGTCTGATGTGCGAGCTTTCAGCGGAGCAATATAGTCTCTGGAGCC 180
DB 121 AGAACCTTAAGTCTGATGTGCGAGCTTTCAGCGGAGCAATATAGTCTCTGGAGCC 180
QY 181 GGTGCTTTTGTAGCGGTTCTACTTTACAATATACCGGTCTTTACGGTTACTGATGCT 240
DB 181 GGTGCTTTTGTAGCGGTTCTACTTTACAATATACCGGTCTTTACGGTTACTGATGCT 240
QY 241 GAGTAAGTCTTGTGATTAAGTAAATTTTTCAGTAACT 300
DB 241 GAGTAAGTCTTGTGATTAAGTAAATTTTTCAGTAACT 300

DB 241 GACGTAAGTGTTCGTGCAATAGATTTAAATTAATTTTTCAGCAGGCTCTTTTTCAGTAACT 300
QY 301 GGTGATATTTTCATAGGTTTCAGTGTAGATACGGGAGAGCTAATAAATTTTCAGTAACT 360
DB 301 GGTGATATTTTCATAGGTTTCAGTGTAGATACGGGAGAGCTAATAAATTTTCAGTAACT 360
QY 361 ATTGATGATGTTTAAACCTTAACTTTTACAGGTACCGGTACTGAGCCTACGGTGCAAT 420
DB 361 ATTGATGATGTTTAAACCTTAACTTTTACAGGTACCGGTACTGAGCCTACGGTGCAAT 420
QY 421 CCTCGTGTGTTTATTCAGAGTGGACAGCTGCTGCTAATAATATACATATACATATAGT 480
DB 421 CCTCGTGTGTTTATTCAGAGTGGACAGCTGCTGCTAATAATATACATATACATATAGT 480
QY 481 AATTAACCTCTAGTGGAGCAATGCGGTTTGAATTTTTCAGATCCAGATGATTA 540
DB 481 AATTAACCTCTAGTGGAGCAATGCGGTTTGAATTTTTCAGATCCAGATGATTA 540
QY 541 GGACCAATAACGCTTTCAGAGGAGTATAGATGGAGGAGTATATAAATGCAATACAGAT 600
DB 541 GGACCAATAACGCTTTCAGAGGAGTATAGATGGAGGAGTATATAAATGCAATACAGAT 600
QY 601 GCTGCCATTAAACGCAATAGTAACTAATCCGCGAGCTCAATTAAGCATTTGGAGCA 660
DB 601 GCTGCCATTAAACGCAATAGTAACTAATCCGCGAGCTCAATTAAGCATTTGGAGCA 660
QY 661 AGTACGCTTTCTTGGAGGCGCAGTTATTAAGCCACTACGACTTAAGTTAAACAAATGCA 720
DB 661 AGTACGCTTTCTTGGAGGCGCAGTTATTAAGCCACTACGACTTAAGTTAAACAAATGCA 720
QY 721 GCGCGGTATTAACACCTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
DB 721 GCGCGGTATTAACACCTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
QY 781 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 781 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
QY 841 AATATAGGTATTAACAAATTTCAATAGCAGCAATAGTGTAGGAGGAGTACGGTACGTTA 900
DB 841 AATATAGGTATTAACAAATTTCAATAGCAGCAATAGTGTAGGAGGAGTACGGTACGTTA 900
QY 901 GGGGAGCGGTTTAAAGCCACTACGACTTAAGTTACGATGAGGAGTACGGTACGTTA 960
DB 901 GGGGAGCGGTTTAAAGCCACTACGACTTAAGTTACGATGAGGAGTACGGTACGTTA 960
QY 961 CTTTAAATGCAATTTTAAACAGGTGCTTGTGATAACACCAACAGCGGTGATATGATG 1020
DB 961 CTTTAAATGCAATTTTAAACAGGTGCTTGTGATAACACCAACAGCGGTGATATGATG 1020
QY 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGGAATATAGTAAATACAAATTC 1080
DB 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGGAATATAGTAAATACAAATTC 1080
QY 1081 TTAGCCACGATAAATATAGGAGCAGGTGTGGCTACCTTTGGATGGAGCGGTTATTAAGCT 1140
DB 1081 TTAGCCACGATAAATATAGGAGCAGGTGTGGCTACCTTTGGATGGAGCGGTTATTAAGCT 1140
QY 1141 ACTACGACTAAGTTTAAACAGATGATGCTCAGTATTAATATTTTACGAAATCTCTGTAGT 1200
DB 1141 ACTACGACTAAGTTTAAACAGATGATGCTCAGTATTAATATTTTACGAAATCTCTGTAGT 1200
QY 1201 ACCGGAGCAATAGATATACCGGTAAATGCAATTAAGGTTGGTAAATCTTTTACCGGAGCA 1260
DB 1201 ACCGGAGCAATAGATATACCGGTAAATGCAATTAAGGTTGGTAAATCTTTTACCGGAGCA 1260
QY 1261 AGTACGTAACCGATATATAGGTACACGCGAGTATTACGAGGTTAGCGTAGGAGCA 1320
DB 1261 AGTACGTAACCGATATATAGGTACACGCGAGTATTACGAGGTTAGCGTAGGAGCA 1320
QY 1321 GGTTCGTCGCAATACAAAGCGGAGTAGTAAAGCGCAATGCAATTAACCTTTTACCGGATA 1380
DB 1321 GGTTCGTCGCAATACAAAGCGGAGTAGTAAAGCGCAATGCAATTAACCTTTTACCGGATA 1380

```

QY 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGGTAACAGGTAGTATAGGTGGTACAGAA 1440
|
|
|
Db 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGGTAACAGGTAGTATAGGTGGTACAGAA 1440
|
|
|
QY 1441 CTCCTGCGCAACAGTGAATATAGGACAGGAATAACATTACAGCGCGGAGGAAAGCCTAGCT 1500
|
|
|
Db 1441 CTCCTGCGCAACAGTGAATATAGGACAGGAATAACATTACAGCGCGGAGGAAAGCCTAGCT 1500
|
|
|
QY 1501 GCGAATAATATAGATTTGGAGCTGGAGTAATTTAGAGTTTAAAGCGTCTGCCGGTAAG 1560
|
|
|
Db 1501 GCGAATAATATAGATTTGGAGCTGGAGTAATTTAGAGTTTAAAGCGTCTGCCGGTAAG 1560
|
|
|
QY 1561 AATTATAACTTAAATCGGAACCTATAGCAACCGGTAAATATGCTACACTTAAATATTAATGCT 1620
|
|
|
Db 1561 AATTATAACTTAAATCGGAACCTATAGCAACCGGTAAATATGCTACACTTAAATATTAATGCT 1620
|
|
|
QY 1621 GCTGTACAGTGAATGCAAAATGATGTAGTATAGGCACAGTTGCACAAATTAACATTCAA 1680
|
|
|
Db 1621 GCTGTACAGTGAATGCAAAATGATGTAGTATAGGCACAGTTGCACAAATTAACATTCAA 1680
|
|
|
QY 1681 AATAATAAGATTTTCTAATAAATGCTAAGAAGCGTGTATGATATATTAACGCTCAG 1740
|
|
|
Db 1681 AATAATAAGATTTTCTAATAAATGCTAAGAAGCGTGTATGATATATTAACGCTCAG 1740
|
|
|
QY 1741 GCGATTAGTTTAAAGGAGCAGCTTCACGCTCTTTCTTAGCTAACGTTAGTCTACAGATG 1800
|
|
|
Db 1741 GCGATTAGTTTAAAGGAGCAGCTTCACGCTCTTTCTTAGCTAACGTTAGTCTACAGATG 1800
|
|
|
QY 1801 ATAGATTATCACTCTTAAATAATCAITTTACCCGGTCTTGCTAACGCTGGTGGTAA 1860
|
|
|
Db 1801 ATAGATTATCACTCTTAAATAATCAITTTACCCGGTCTTGCTAACGCTGGTGGTAA 1860
|
|
|
QY 1861 TTCTTTTGTAGTCCAAACAACTTATGACCGTACAGGTGACGTGAGCTAATAACAAATAG 1920
|
|
|
Db 1861 TTCTTTTGTAGTCCAAACAACTTATGACCGTACAGGTGACGTGAGCTAATAACAAATAG 1920
|
|
|

```

RESULT 2

US-09-800-065-3

; Sequence 3, Application US/098000065

; GENERAL INFORMATION:

; APPLICANT: Bouyer, Donald H

; APPLICANT: Croquet-Valdes, Patricia

; APPLICANT: Stenos, John

; APPLICANT: Walker, David H

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR APPLICATION NUMBER: US 60/187,323

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 5513

; TYPE: DNA

; ORGANISM: Rickettsia felis

US-09-800-065-3

Query Match 100.0%; Score 1920; DB 31; Length 5513;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGCGCAATATTTCTCTAAATATTTTCAAAAGCAATTCAAAAGTCTTAAACTGCT 60
|
|
|
Db 478 ATGCGCAATATTTCTCTAAATATTTTCAAAAGCAATTCAAAAGTCTTAAACTGCT 537
|
|
|
QY 61 TTATTACCACTCAACCGCAGCGATAATGCTAACCGGTAGTGGAGTCCCTTGGTGTGCA 120
|
|
|
Db 538 TTATTACCACTCAACCGCAGCGATAATGCTAACCGGTAGTGGAGTCCCTTGGTGTGCA 597
|
|
|
QY 121 AGAACCGTAACCTGCTGATGGTGCAGAGCTTTCAGCCGGAACAAATATAGTCTCTGGAGCC 180
|
|
|

```

```

|
|
|
Db 598 AGAACCGTAACCTGCTGATGGTGCAGAGCTTTCAGCCGGAACAAATATAGTCTCTGGAGCC 657
|
|
|
QY 181 GGTGCTTTTGTAGCGGGTCTACTTACAATATACGGTGTCTTTTACGGTACTGATGCT 240
|
|
|
Db 658 GGTGCTTTTGTAGCGGGTCTACTTACAATATACGGTGTCTTTTACGGTACTGATGCT 717
|
|
|
QY 241 GACGTAAGTGTTCGTCGATTAGATTAAATAATTTTGCAGCAGGCTCTTTTTCAGTAAC 300
|
|
|
Db 718 GACGTAAGTGTTCGTCGATTAGATTAAATAATTTTGCAGCAGGCTCTTTTTCAGTAAC 777
|
|
|
QY 301 GGTGATATTTCAATAGTTCAGTGTAGATACGGGAGGAGCTAAATAAATTCGAGTTAAT 360
|
|
|
Db 778 GGTGATATTTCAATAGTTCAGTGTAGATACGGGAGGAGCTAAATAAATTCGAGTTAAT 837
|
|
|
QY 361 ATTGATGATGGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 420
|
|
|
Db 838 ATTGATGATGGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 897
|
|
|
QY 421 CCTGGCTTCTTATCCAAAGGTGACAAGCTGCTGCTTAATAATACATATATCTGCTTAGCT 480
|
|
|
Db 898 CCTGGCTTCTTATCCAAAGGTGACAAGCTGCTGCTTAATAATACATATATCTGCTTAGCT 957
|
|
|
QY 481 AATATAACTCTTAGTGTGAGCGAATGCCGTTTCACTTATTCAGTCCAGATCCAGATGATTA 540
|
|
|
Db 958 AATATAACTCTTAGTGTGAGCGAATGCCGTTTCACTTATTCAGTCCAGATCCAGATGATTA 1017
|
|
|
QY 541 GGACCAATAAACCGTTTCAGGAAATATAGATGGAGGAGGTATATACTGACATACAGAT 600
|
|
|
Db 1018 GGACCAATAAACCGTTTCAGGAAATATAGATGGAGGAGGTATATACTGACATACAGAT 1077
|
|
|
QY 601 GCTGCCATTAAACGGAACAATAGTAACTACTTCCGGCAGCTCAAAATAAGCATTTGAGCA 660
|
|
|
Db 1078 GCTGCCATTAAACGGAACAATAGTAACTACTTCCGGCAGCTCAAAATAAGCATTTGAGCA 1137
|
|
|
QY 661 AGTACGCTTTCTCTTGGAGGGGCGAGTTATTAAGCCACTACGACTAAGTTAAACAAATGA 720
|
|
|
Db 1138 AGTACGCTTTCTCTTGGAGGGGCGAGTTATTAAGCCACTACGACTAAGTTAAACAAATGA 1197
|
|
|
QY 721 GCGCCGGTATTACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
|
|
|
Db 1198 GCGCCGGTATTACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1257
|
|
|
QY 781 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
|
|
|
Db 1258 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
|
|
|
QY 841 AATATAGTAAATACAAATTCATTAGCGACAATAAGTGTAGGAGCAGGTACGGCTAGCTTA 900
|
|
|
Db 1318 AATATAGTAAATACAAATTCATTAGCGACAATAAGTGTAGGAGCAGGTACGGCTAGCTTA 1377
|
|
|
QY 901 GGGGAGCGGTTTAAAGCCACTACGACTAAGTTAAAGCCACTACGACTAAGTTAAAGCC 960
|
|
|
Db 1378 GGGGAGCGGTTTAAAGCCACTACGACTAAGTTAAAGCCACTACGACTAAGTTAAAGCC 1437
|
|
|
QY 961 CTTTACAAATGCACTATTAAACAGTGTGTTGATAACACACACACACACACACACACAC 1020
|
|
|
Db 1438 CTTTACAAATGCACTATTAAACAGTGTGTTGATAACACACACACACACACACACACAC 1497
|
|
|
QY 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGGAATATAGTAAATACAAATTC 1080
|
|
|
Db 1498 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACCGGGAATATAGTAAATACAAATTC 1557
|
|
|
QY 1081 TTAGCCACCAATAATATAGGAGCAGGTGCTGCTACCTTGGATGGAGCGGTATTAAAGCT 1140
|
|
|
Db 1558 TTAGCCACCAATAATATAGGAGCAGGTGCTGCTACCTTGGATGGAGCGGTATTAAAGCT 1617
|
|
|
QY 1141 ACTTACGACTAAGTTAAACAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
|
|
|
Db 1618 ACTTACGACTAAGTTAAACAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677
|
|
|
QY 1201 ACCGGAGCAATAGATATACCGGTAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1260
|
|
|

```

```
Db 1678 ACCGGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGGTAACTTTACCGGAGCA 1737
Qy 1261 AGTACGGTAACCGATAATATAGGTAAACAGCGCAGTATTAGCAGAGGTAAAGCGTAGGAGCA 1320
Db 1738 AGTACGGTAACCGATAATATAGGTAAACAGCGCAGTATTAGCAGAGGTAAAGCGTAGGAGCA 1797
Qy 1321 GGTTCGCTGCAATACACAGCGGAGTAGTAAAGCGAATGCATAACTTACGGATAT 1380
Db 1798 GGTTCGCTGCAATACACAGCGGAGTAGTAAAGCGAATGCATAACTTACGGATAT 1857
Qy 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTAGCGTAACAGGTAGTATAGGTGGTACAGAA 1440
Db 1858 GCGTCAGTAGTAACATTTACCGGTGATAGTAGCGTAACAGGTAGTATAGGTGGTACAGAA 1917
Qy 1441 CTCCTCGCAACAGTGAATATAGGAGCAGGAATAACATTTACGAGCGGAGGAGCCTAGCT 1500
Db 1918 CTCCTCGCAACAGTGAATATAGGAGCAGGAATAACATTTACGAGCGGAGGAGCCTAGCT 1977
Qy 1501 GCGAATAATATAGATTTTGGAGCTCGGAGTAATTTAGAGTTTACGGTCCGCGGTAG 1560
Db 1978 GCGAATAATATAGATTTTGGAGCTCGGAGTAATTTAGAGTTTACGGTCCGCGGTAG 2037
Qy 1561 AATTATACTTAATCGGAACATAGCAACCGTAATATGCTACACCTTAATATTAATGCT 1620
Db 2038 AATTATACTTAATCGGAACATAGCAACCGTAATATGCTACACCTTAATATTAATGCT 2097
Qy 1621 GCTGTCAGTATGCAATATGCTAGTATAGGAGCAGGTGACACAGTTGACAAATTAACATTCAA 1680
Db 2098 GCTGTCAGTATGCAATATGCTAGTATAGGAGCAGGTGACACAGTTGACAAATTAACATTCAA 2157
Qy 1681 AATAATAGATTTTGTAAATAGCTAAGAACGCTGATGTTGATATATATAGAGCCTCAG 1740
Db 2158 AATAATAGATTTTGTAAATAGCTAAGAACGCTGATGTTGATATATATAGAGCCTCAG 2217
Qy 1741 GCGATTAGTTTAAAGGAGCAGCTCAGCTCTTTCTTAGCTAACGTTAGTCTACAGATG 1800
Db 2218 GCGATTAGTTTAAAGGAGCAGCTCAGCTCTTTCTTAGCTAACGTTAGTCTACAGATG 2277
Qy 1801 ATAGAGTTATCACTTTAAATAATCATTTACCGGCTCTTGTGTAACGGTGGTGGTGAATAA 1860
Db 2278 ATAGAGTTATCACTTTAAATAATCATTTACCGGCTCTTGTGTAACGGTGGTGGTGAATAA 2337
Qy 1861 TTCCTTTGAGTCCCAACAACTTATGACCGTACAGGTGAGGTGGAGCTAAACAAATAG 1920
Db 2338 TTCCTTTGAGTCCCAACAACTTATGACCGTACAGGTGAGGTGGAGCTAAACAAATAG 2397
```

RESULT 3

```
US-08-029-329-1
; Sequence 1, Application US/08029329
; GENERAL INFORMATION:
; APPLICANT: SUMNER W., JOHN
; APPLICANT: ANDERSON E., BURT
; APPLICANT: PELLET E., PHILIP
; APPLICANT: SANCHEZ-MARTINEZ, DEMETRIO
; TITLE OF INVENTION: "BACULOVIRUS EXPRESSED 190 kDa ANTIGENS
; TITLE OF INVENTION: OF RICKETTSIA RICKETTSII AND RELATED METHODS"
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/029,329
; FILING DATE: 19930309
```

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7088 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70..6816
; US-08-029-329-1
```

```
Query Match 31.1%; Score 597.6; DB 4; Length 7088;
Best Local Similarity 63.9%; Pred. No. 7.1e-143;
Matches 982; Conservative 0; Mismatches 518; Indels 36; Gaps 4;

Qy 1 ATGCGGAATATTTCTTAAATATTTCAAAAAGCAATTTCAAAAGGTCTTTAAACTGCT 60
Db 70 ATGCGGAATATTTCTTAAATATTTCAAAAAGCAATTTCAAAAGGTCTTTAAAGCCCT 129
Qy 61 TTATTACCACCTCAACCGCAGCATATGCTAACCGGTAGTGCAGTCTTGGTGTGCTCA 120
Db 130 TTATTACCACCTCAACCGCAGCATATGCTAGTAGTGGGGGACCTCGGTGTGCT 189
Qy 121 AGAACCGTAACCTGATGTGTCAGAGCTTGCAGCGCGCAACAAATATAGGTCTTGGAGCC 180
Db 190 ACAGGTGTATTGCTACTAATAATATGTCAGCATTTAGTATATATGTTGCAATAATAAT 249
Qy 181 GGTGCTTTTGTAGCGGTCTTACTTTACATATATACCGGTGCTTTTACGGTTACTGATCT 240
Db 250 TGGAAATGAGATAACGGCTGCAGGGGTAGCTAATGCTACTCTGCTGCGGTCTCTCAAAAC 309
Qy 241 GACGTAAGTGTTCGTGTCATTAGATTTAAATAATTTTCAGCAGGTCTTTTTCAGTAAT 300
Db 310 AATTGGCAATTTACTTACGGTGTGATTTACTGCTACTGCAGATGACGCGGATCGTATT 369
Qy 301 GGTGATATTTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTAA 360
Db 370 ATTAAGCTATATAATGTTGCGGTACTACTCCCGGTAGTCTAAATATATTAATAATACT 429
Qy 361 ATGATG-----ATGTTTAACTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 399
Db 430 GTCGTTGGTTCGATTTAAGCAAGGTAACTTTGTCCTGTTACTTCTTAATGCGGCAAA 489
Qy 400 ACTGCAGCTACGGTGCAATCTCGCTGTTGTTATTCAGGTGTCAGAGCTGCTGCTTAAT 459
Db 490 AGCTTAACTTAAATGTTAAATGCTGTTGCTGCAATCATGTTTGTATGCGCTGCTGCC 549
Qy 460 AATACATATATCTTCTTGTAGTATATATCTTCTAGGTGAGCGAATGCGGTGTTGACTATT 519
Db 550 GATAATTTATACAGTTTGTAGGAATATAGCTTTAGGGGAGCGAATGCTGCACATTAATAT 608
Qy 520 GCTTCAGATCCAGATGTTATGAGCAACCAATTAACCGTTGCAGGAATATATAGATGAGGAGGT 579
Db 609 -----ACAATCTCGAGCTCGTCAAAAGATAACACTTTGCAAGGAATATATAGATGAGGAGGT 663
Qy 580 ATAATACTGACAATACAGATGCTGCAATTTAAGCAACAAATAGTATATCTAATCGGCA 639
Db 664 ATAATACTGACAATACAGATGCTGCAATTTAAGCAACAAATAGTATATCTAATCGGCA 723
Qy 640 GCTCAATAAGCAATTTGAGCAAGTACGCTTTCTTCTTGGAGGGCAGTATTAAGGCACT 699
Db 724 GCAACAGTGAATGTAGGAGCAGGTACAGCCACCTTAGGGGAGCGGTATTAAGGCACT 783
Qy 700 ACAGCTAAGTTAACAATGCAGCGCGCGGTATTAACACTTTAACAATGCAGATGATTA 759
```



```

RESULT 8
US-60-233-468-1986/c
: Sequence 1986, Application US/60233468
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: APPLICANT: Valdes, Ana
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with ADME Genes
: FILE REFERENCE: GX-0013-2 P
: CURRENT APPLICATION NUMBER: US/60/233,468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 2488
: SOFTWARE: PERL Program
: SEQ ID NO 1986
: LENGTH: 7814
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: GBI:AC024253_3_0000004
US-60-233-468-1986

```

Query Match	4.08;	Score 76.8;	DB 67;	Length 7814;	
Best Local Similarity	43.08;	Pred. No. 2.2e-08;			
Matches 433;	Conservative	0;	Mismatches 572;	Indels 3;	Gaps 1;
QY	509	GTTFGACTATTGCTTCAGATCCAGATGTTATAGACCAATAAACGCTTGACGAATATAG	568		
Db	3851	GATTAAACAGCAGCATCAATAGCCACAGCAACAGCAGCAGTAGTACCAACAGAAATCG	3792		
QY	569	ATGGAGGAGGTATTAATAACTGTACAATACAGATGCTGCCATTAAACGGAAACAATAGTGAATA	628		
Db	3791	CAGTAATAAACAGTAGTACGAGCAGCAGCAGCAATCATTTACCATTAGTACCAACAACAGTAG	3732		
QY	629	CTAATCCGGCAGTCGAATAGCATTTGAGCAAGTAGCGTTCTCTTGGAGGGCGCAGTTA	688		
Db	3731	CAGCAACACCCAGCAGCAATAGTAGCAGCAATAGCAACAGTAATAGTAGCAACCAACAGCAA	3672		
QY	689	TTAAAGCCACTACGACTAAGTTAACTAAATGCGAGCCGCGGTATTAACTCTTACAAATGCAA	748		
Db	3671	TAGCAACAGTAGCAGCATCAGCAACAACAGCAGCAGCAACAATAAGACGACGAC	3612		
QY	749	ATGCAGTATTAAACAGTGTCTTGTATACACCACAGCGCGGTGATGTAGGTGTCATTAA	808		
Db	3611	CAGCAACAACCAACAGTAGTACGACGAGCAACAGCAACAACAAATAACAAAAGCAACGATAG	3552		
QY	809	ATTTAAACGGGCGCTTGAGTCAAGTTAACTGTGAAATATAGTTAATAACAAATTCATTAGCGA	868		
Db	3551	CAACAGTATCAGCAACAGCAACAGCATCAGCAACAACAGCAACAATAGCAGCAGTAAAA	3492		
QY	869	CAATAAGTGTAGGACAGGTACGGCTAGTTAGGGGAGCGGTTATTAAAGCCACTACGA	928		
Db	3491	CAGCAACAGCAGCAGCAACACAGCAACAGCAGCAGCAACAATAAGAGCAGCAACAACAG	3432		
QY	929	CTAAGTTAAACGAATGCAGCGTCGGTATTAAACACTTACAAATGCAGTATTAAACAGGTGCTG	988		
Db	3431	CAACAGTAGCAGCAGCAACACAGCAGCAACAACAAGCAACAATAAGCAGCAGTAT	3372		
QY	989	TTGATACACCACAGCGCGGTGATTAATGTAGTGTGCTGTAATCTTAAGTGGAGC---ATTGA	1045		
Db	3371	CAGCAACAGCAACAGCAATAGCAACAAGCATCAGCAACAGCAATAAGCATCATCAAAATTAA	3312		
QY	1046	GTCAAAGTAAACCGGAATATAGTGAATAACAATTCATTAGCCACCATTAATAATAGGACGAC	1105		
Db	3311	CAACAGCAACAGCAGCAGTAGCAACAGAAATCAGATTAACAACAGTAGCAGCAGCAGCAG	3252		
QY	1106	GTGTGGCTACCTTGGATGGAGCGGTTATTAAAGCTACTACGACTTAAGTTAACAGATGATG	1165		
Db	3251	CAATAAATTACCATAGTAGCAACAACAGTAGCAGCAACAACAGCAGCAGTAGCAGCAGCA	3192		

```

RESULT 9
US-60-313-371-1986/c
; Sequence 1986, Application US/60313371
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with ADME Genes
; FILE REFERENCE: GX-0013-5 P
; CURRENT APPLICATION NUMBER: US/60/313,371
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 1986
; LENGTH: 7814
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GBI:AC024253_3_000004
US-60-313-371-1986

```

Query Match	4.08;	Score 76.8;	DB 75;	Length 7814;
Best Local Similarity	43.08;	Pred. No. 2.2e-08;		
Matches 433;	Conservative 0;	Mismatches 572;	Indels 3;	Gaps 1;
QY 509	GTITTCACCTATTGCTTCAGATCCAGATGTAITTAGGACCAATAACGCTTTCAGGAAATATAG	568		
Db 3851	GATTAACAGCAGCATCAATAGCCACACAGCAACAGCAGCAGTAGTAGCAACAGAAATCG	3792		
QY 569	ATGGAGGAGGTATTAATAACTGACAATACAGATGCTGCCATTAAACGGCAACATAGGTAA	628		
Db 3791	CAGTAATAACAGTAGCAGCAGCAGCAGCAATCATTACCATAGTAGCAACACACAGTAG	3732		
QY 629	CTAATCCGGCAGCTCAAATAGCAATTGGAGCAAGTAGCTTTCTCTTTGGAGGGGCAGTTA	688		
Db 3731	CAGCAACACCGCAGCAATAGTAGCAGCAATAGCAACAGTAATAGTAGCAGCAACACGAA	3672		
QY 689	TTAAGCCACTACGACTAAGTTTACAAATGAGCGCCGGTATTAAACACTTTACAANTGCAG	748		
Db 3671	TAGCAACAGTAGTAGCAGCATCAGCAACACAGCAACAGCAGCAGCAACAATATAGACGAGCAG	3612		
QY 749	ATGCAGTATTTAACAGGTGCTTTCATTAACACCAGCGGGTGATGATCTAGGTGCTCTTAA	808		
Db 3611	CAGCAACCAACCAACAGTAGCAGCAGCAGCAACAGCAACAATTAACAAAAGCAGCATAG	3552		
QY 809	ATTTAAACGGAGCGTTGAGTCAAGTAACTGGAATAATAGGTAATACAAATTCATTAGCGA	868		


```

; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
; FILE REFERENCE: GX-0013-2 P
; CURRENT APPLICATION NUMBER: US/60/233,468
; CURRENT FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 2488
; SOFTWARE: PERL Program
; SEQ ID NO 1976
; LENGTH: 186935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:AC022322_3
US-60-233-468-1976

      8392 CTGCTACTGCTGCTGGTGTGCTGCTACTGTTGTTGCTACTATGATGTAATATTGCTGCTG 8333
      155  CCGGAACAAATATAGTCTCCTGGAGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
      8332 CTGCTGCTACTGTTGTTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8273
      215  CCGGTAACAAATATAGTCTCCTGGAGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 274
      8272 GTGATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8213
      275  TTGCAGCAGCTCTTTTTCAGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 334
      8152 TTGCTGCTCTTATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8093
      392  GTACCGGTACTGCGAGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
      8092 CTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8033
      452  CTGCTAAATATACATATCTGCTTTAGGTAATATACTCTAGGTGAGGCGAATGCGCGTT 511
      8032 TTGCTTTGTTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7973
      512  TGACTATTGCTTCCAGATCCAGATGATATAGGACCAATAGCGTTGCGAGGAATATAGATG 571
      7972 CTCTTATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7913
      572  GAGGAGTATAATACTGACAAATACAGATGCTGCCATTACCGAACAATAAGGTAATACTA 631
      7912 TTGCTGCTACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7853
      632  ATCCGCGAGCTCAAAATAGCAAGTACGCTTTCTCTTGGAGGGCAGTTAATA 691
      7852 TTGCTGCTACTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7793
      692  AAGCCACTAGGACTAAGTTAACAAATGACGCGCGGTTATTAACACTTACAA -----ATG 745
      7792 CTGTTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7493
      746  CAAATGAGTTAAGCAATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 985
      7452 CTACTGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7493
      986  CTGTTGATACACACAGCGGTTGATATAGTAG 1018
      7492 CTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7460
```

RESULT 12

```

US-60-233-468-1976/c
; Sequence 1976, Application US/60233468
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; APPLICANT: Valdes, Ana
```

```

Query Match      3.9%; Score 74.2; DB 67; Length 186935;
Best Local Similarity 44.1%; Pred. No. 3.4e-07;
Matches 411; Conservative 0; Mismatches 513; Indels 9; Gaps 2;

      95  CCGGTAGTGGAGTCCCTTGTGCTGCAAGAACCGTAACCTGCTGATGCTGAGAGCTTGCGAG 154
      8392 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8333
      155  CCGGAACAAATATAGTCTCCTGGAGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
      8332 CTGCTGCTACTGTTGTTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8273
      215  CCGGTGCTTTTACGGTTACTGATGCTGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 274
      8272 GTGATGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8213
      275  TTGCAGCAGCTCTTTTTCAGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 334
      8212 TTGCTAATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8153
      335  GAGGAGTAAATAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 391
      8152 TTGCTGCTCTTATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8093
      392  GTACCGGTACTGCGAGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
      8092 CTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8033
      452  CTGCTAAATATACATATCTGCTTTAGGTAATATACTCTAGGTGAGGCGAATGCGCGTT 511
      8032 TTGCTTTGTTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7973
      512  TGACTATTGCTTCCAGATCCAGATGATATAGGACCAATAGCGTTGCGAGGAATATAGATG 571
      7972 CTCTTATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7913
      572  GAGGAGTATAATACTGACAAATACAGATGCTGCCATTACCGAACAATAAGGTAATACTA 631
      7912 TTGCTGCTACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7853
      632  ATCCGCGAGCTCAAAATAGCAAGTACGCTTTCTCTTGGAGGGCAGTTAATA 691
      7852 TTGTTGCTACTATGGAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7793
      692  AAGCCACTAGGACTAAGTTAACAAATGACGCGCGGTTATTAACACTTACAA -----ATG 745
      7792 CTGTTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7493
      746  CAAATGAGTTAAGCAAGTGTGTAACACACAGCGGTTGATGATGAGGTAAGGTAAGTGTCT 805
      7732 CTATTGCTGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7673
      806  TAAATTTAAGCGAGCTGAGTCAAGTAACTGGAATATAGGTAATAACAAATTCATTAAG 865
      7672 CTTTTGTATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7613
      866  CGACAATAAGTGTAGGACGAGGTACGGCTACGTTAGGGGAGCGGTTATTAAGCCACTA 925
```

```

Db      7612  TTGCTGCTGCTGCTGCTGTTGTTACTGCTGCTATTTGCTGTTATTTGTTGCTATGCTG  7553
QY      926  CGACTAAGTAAACGAATGCAGCGCTCGGTATTAAACACTTACAAGTCAGTATTAAACAGGTG  985
Db      7552  CTACTGTTGCTATTGCTGTTGCTGCTACTATTACTGTTGCTATTGCTGCTGCTACTGCTG  7493
QY      986  CTGTTGATAACACACACAGCGCGTGATAATGTAG  1018
Db      7492  CTGGTGTGCTGCTACTGTTGTTGCTACTATGG  7460

RESULT 13
US-60-313-371-1976/c
; Sequence 1976, Application US/60313371
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
; FILE REFERENCE: GX-0013-5 P
; CURRENT APPLICATION NUMBER: US/60/313,371
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 1976
; LENGTH: 186935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:AC022322_3
US-60-313-371-1976

```

Query Match	3.98;	Score 74.2;	DB 75;	Length 186935;
Best Local Similarity	44.1%;	Pred. No. 3.4e-07;		
Matches 411;	Conservative 0;	Mismatches 513;	Indels 9;	Gaps 2;
QY	95	CCGGTAGGAGTCCCTGGCTGCGACGCGGCTTTGTAGCGGGTCTACTTTACAATATA	154	
Db	8392	CTGCTACTGCTCGGGTGTGCTGCTACTGTGTGCTACTATGGTAATATTGCTGCTG	8333	
QY	155	CCGGAACAAATATAGGTCTCGAGCGGCTTTGTAGCGGGTCTACTTTACAATATA	214	
Db	8332	CTGCTGCTACTGTTGTACTGTATTCGTGCTACTGCTGCTGCTGTTGTTAAT	8273	
QY	215	CCGGTGTCTTTACGGTACTGATGCTGACGPAAGTGTGTCGCAATTAGATTTAAATAAT	274	
Db	8272	GTGATGTATTGCTGTGCTGATGCTTTGTGCTATTGCTGTGCTGCTGCTGCTGCTG	8213	
QY	275	TTGCAGCAGGCTTTTTTTCAGTAACACTGGTGATATTTCAATTAGGTTTCAGTGTAGATACGG	334	
Db	8212	TTGCTATTGTGCTTTGTTGCTGCTGCTGTTGTGCTGCTGCTACTGTTGCTGTTG	8153	
QY	335	GAGGAGCTAATAACTTGCAGCTTAATATTGA--TGATGGTTTAAACCTTAACTTTAACAG	391	
Db	8152	TTGCTGCTCTATTGTTGCTGCTGCTGCTGCTGTTGCTGCTGCTGCTGCTGTTTTTA	8093	
QY	392	GTACCGGTACTGACGCTACGGTGCAAATCCTCGGTTGTTATCCAAAGTGGACAAGCTG	451	
Db	8092	CTGCTGCTATTGTTGCTGTTGTTGCTGATGCTGTTGCTGCTGCTGCTGCTGCTGCTG	8033	
QY	452	CTGCTAATAATACATACTGCTTTAGTAAATAAATCTAGGTGGAGCGGAATGCCGGTT	511	
Db	8032	TTGCTTTTGTATTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7973	
QY	512	TGACTATTGCTTCAGATCCAGATGTATTAGGACCAATAAGCGTTGCGAGGAATAATAGATG	571	
Db	7972	CTCTTATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7913	
QY	572	GAGGAGGTAAATAACTGACAAATACAGATGCTGCCATTACGGNACAATAAGGTAAATACTA	631	
Db	7912	TTGCTGCTACTATTACGTTGCTATTGCTGCTACTATTGCTGCTGCTGCTGCTGCTGCTG	7853	

QY	632	ATCCGGCAGCTCAAAATAAGCATTTGGACGAAGTACGCTTTCTCTTGAGGGCGAGTTATTA	691
Db			
Db	7852	TGTGTGCTACTATAGTAAATGATTGCTGCTGCTGCTGCTACTGTTATTACTGCGAATT	7793
QY	692	AAGCCACTACCACHTAAGTTTAAACAATTCACAGCGCGGTATTACACTTACAA-----ATG	745
Db			
Db	7792	CTGTGTCTACTGCCTGCTGTTGCTGTGCTGTGGCTATTGCTGCTGTTAATTGTGATA	7733
QY	746	CAAAATCAGTATTAAACAGGTGCTTTTGATAAACACACAGCGCGGTGATGATGTAGTGCT	805
Db			
Db	7732	CTATTGCTGTGCTGATGCTGCTGTTGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7673
QY	806	TAAATTTAAACGGAGCGTTGAGTCAAAGTAACTGGAATAATAGTAAATACAAATTCATTAG	865
Db			
Db	7672	CTTTTGTATTGTTGCTCTTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTATTATTG	7613
QY	866	CGACAATAAGTGTAGGAGCAGGTACCGCTACGTTAGGGGGAGCGGTTATTAAAGCCACTA	925
Db			
Db	7612	TGTGCTGCTGTTGCTGTTGTTTACTGCTGCTATTGTTGCTGTTTATGTTGCTGATGCTG	7553
QY	926	CGACTAAGTTTAAACGATCGACGCTCGGTATTAACTATTAACTATCAAAATGCAGTATTAA	985
Db			
Db	7552	CTACTGTTGCTATTGCTGTTGCTGCTACTATTACTGTTGCTATTGCTGCTGCTACTGCTG	7493
QY	986	CTGTTGATACACACAGCGCGGTGATAATGTAG	1018
Db			
Db	7492	CTGTGTTGCTGCTACTGTTGTTGCTACTATGG	7460
RESULT 14			
US-60-360-207-31796/c			
; Sequence 31796, Application US/60360207			
; GENERAL INFORMATION:			
; APPLICANT: VENTER, J. Craig			
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF			
; FILE REFERENCE: CL001321			
; CURRENT APPLICATION NUMBER: US/60/360,207			
; CURRENT FILING DATE: 2002-03-01			
; NUMBER OF SEQ ID NOS: 47235			
; SEQ ID NO 31796			
; LENGTH: 961			
; TYPE: DNA			
; ORGANISM: HUMAN			
US-60-360-207-31796			
Query Match 3.7%; Score 71.2; DB 80; Length 961;			
Best Local Similarity 46.4%; Pred. No. 2.7e-07;			
Matches 232; Conservative 0; Mismatches 268; Indels 0; Gaps			
QY	111	TGTTGCTGCNAGAACCGTAACCTGATGGTGTCAGAGCTTGCAGCGCGGACACAATATAGG	170
Db			
Db	677	TGCTGTGTTGTTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	618
QY	171	TCCTGGACGGGTGCTTTTGATACGGGGTCTACTTTACAATATACCGGTGCTTTTAGCGT	230
Db			
Db	617	TGTTGTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	558
QY	231	TACTGATGCTGACGTAAGTGTTCGTCATTAAGATTTAAATTAATTTTCACAGAGCTTTTT	290
Db			
Db	557	TGCTGCTACTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	498
QY	291	TTCACTAACTGGTGATATTTCAATAGGTTCAAGTGTAGTGTAGATACGGGAGAGCTAATAA	350
Db			
Db	497	TGCTGCTGTTGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	438
QY	351	TGCAGTTAATATTGATGATGGTTTAAACCTTAACCTTTAAACAGGTACCGGTACTCGAC	410
Db			
Db	437	TGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	378
QY	411	CGGTGCAAACTCGGTGTTTATTCGAAGGTGGACAGCTGCTGCTTAATAATACATATAC	470
Db			
Db	377	TGTTGTACTGCTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	318

Qy 471 TGC^{TT}TAGGTAATAA^{CT}CTAGGTGGAGCGAA^{TGCCGGTT}GACTATTGCTTCAGATCC 530

D _b	317	TGTTGTGCTGCTGTTGTTGTTGCTGCTGTTGGTGGTGCTACTGTTGTTGCTGCTGC	258
Q _y	531	AGATGTAATTAGGACCAATAACCGTTTCAGGAATAATAGATGGAGGAGGTATAAATACTC	590
D _b	257	TGCTGTTGTTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC	198
Q _y	591	CAATACAGATGCTGCCATTA	610
D _b	197	TTCTGTGCTGCTGCTGCTGTTA	178

RESULT 2

```

US-09-949-016-4495
; Sequence 4495, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4495
; LENGTH: 4184
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4495

```

	Query Match	3.1%	Score 59.6;	DB 7;	Length 4184;
	Best Local Similarity	44.0%;	Pred. No. 0.00012;		
	Matches 407;	Conservative 0;	Mismatches 504;	Indels 15;	Gaps 3;
QY	576	AGGTATAATAACTGACAATACAGATGCTGCCATTAAACGGGRACAATAGTAATACTAATCC	635		
DB					
QY	2026	AGTGACAGTAAAGTCAGACAGAGTCAGACGACAGCAGTAGTACAGTATGACACACAGTAT	2085		
DB					
QY	636	GGCAGCTCAAAATAGCAATTGGAGCAAGTAGCGTTCTCTGGAGGGCGCATTTATAAGC	695		
DB					
QY	2086	AGCAGCGACAGCAGCAATAGCAGTAACAGCAGTGATAGTAGTGACACAGTGATAGCAGT	2145		
DB					
QY	696	CACACACAGTAAAGTTACAAATGACGGCGCGTATTAAACACTTACAAATGCAATAGCAGT	755		
DB					
QY	2146	GACAGCAGCAGTAGCAGTGACAGCAGCAGTAGCAGTGACAGCAGCAACAGCAGTGATAGT	2205		
DB					
QY	756	ATTAACAGGTGCTTGTATTAACACACAGCGGGTGATGTAGGTGCTCTAAATTTAAA	815		
DB					
QY	2206	AGTGACAGTATAGTGACACAGCAATAGCAGTGAGACAGCAGTGATAGTAGTGACACAGTAT	2265		
DB					
QY	816	CGGAGCGGTTGAGTCAAGTAATCGAAATATAGGTAATACAAATTCATTAGCGACAATAAG	875		
DB					
QY	2266	AGTGACAGCAGTGATAGTAGTGACAGCAGTAATAGTAACAGCAGCGATAGTGACAGCAGC	2325		
DB					
QY	876	TGTAGGAGCAGGTACGGCTACGTTAGGGGAGCGGTTATTAAAGCCACTACGACTAAGTT	935		
DB					
QY	2326	AACAGCAGCATAGCAGTGACAGCAGTGATAGCAGTGACAGCAGCAGCAGTGACAGT	2385		
DB					
QY	936	AACGAATGACGCGTCGGTATTACACTTACAAATGACAGTATTACAGCGTGCTGT-----	989		
DB					
QY	2386	AGCGATAGCAGTGACAGCAGCAACAGCAGTGACAGCAGTGATAGCAGTGACAGCAGTGAT	2445		
DB					
QY	990	-----TGATAACACACAGCGGTGATAATGTAGGTGTCGTAATCTAATCTAAGTG--CAGCATTT	1043		
DB					
QY	2446	AGTAGTGACAGCAGCAACAGCAGTGATAGCAACGACAGCAGCAATACGACAGTGACAGCAGT	2505		
DB					
QY	1044	GAGTCAAGTAACCGGGAATATAGGTAAATACAAATTCATTAGCCACCGATAAATATAGGAGC	1103		

RESULT 3

```

US-10-170-235-27462
; Sequence 27462, Application US/10170235
; GENERAL INFORMATION:
;
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC
; TITLE OF INVENTION: TRANSCRIPTS, FOR DET
; FILE REFERENCE: CLO01380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 27462
; LENGTH: 4184
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-27462

```

[illegible]

QY	756	ATTAAACAGGTGCTGTGTGATAACACACACAGCGCGTGATGATGTAGTGTCTTTAAATTTAAA	815
Db	2206	AGTGCAGTAGTGACAGCAGCAATAGCAGTGACAGCAGTGATAGTAGTGACAGCAGTGAT	2265
QY	816	CGGAGCGTTGAGTCAAGTAACCTGGAATATAGGTAATACAATTCATTAGCGCAATTAAG	875
Db	2266	AGTGCAGCAGGTGATAGTAGTGACAGCAGTAAATAGTAACAGCAGCAGTGATAGTGACAGCAGC	2325

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ANALYSIS OF GENE EXPRESSION IN BT 474
; CURRENT APPLICATION NUMBER: US/10/203.138A
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15438
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 7146
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
; FEATURE:
; OTHER INFORMATION: NT HIT: U01287.1, EVALUE 1.30e+00
; US-10-203-138A-7146

Query Match 3.18; Score 59.2; DB 9; Length 510;
Best Local Similarity 46.18; Pred. No. 9.5e-05;
Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;
QY 104 GAGTCCTTGGTGGCGAAGAACCGTACTGCTGATGCTGAGAGCTTGCAGCCGGAACAA 163
DB 510 GTGTGATGCTGATGCTGAGGTACCGTACCGTGTGATGCTGAGGTGCTGAAAAAG 451
QY 164 ATATAGGTCTCGAGCGGCTCTTTTGTAGGGGTCTTACTTTACAATATACCGTCTT 223
DB 450 GTGTGATGCTGAGGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 391
QY 224 TTACGGTTACTGCTGACGTAAGTGTCTGCTGATGATTTAAATAATTTTCAGCAG 283
DB 390 ATGAGGTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 331
QY 284 GTCTTTTTCAGTAACCTGCTGATATTTTCATGATGCTGCTGCTGCTGCTGCTGCTG 343
DB 330 GTGTGATGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271
QY 344 ATAACTTGCAGTTAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
DB 270 CAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 211
QY 404 CAGCTTACGGTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
DB 210 CAAGTGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 151
QY 464 CATATCTGCTTTAGTGAATATATCTAGTGGAGCGAATGCGGTTTGACTATTCCTT 523
DB 150 TTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91
QY 524 CAGATCCAGATG 535
DB 90 CTGATACAGTG 79

RESULT 6
US-10-144-771-31796

; Sequence 31796, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 31796
; LENGTH: 961
; TYPE: DNA
; ORGANISM: HUMAN
; US-10-144-771-31796
Query Match 3.08; Score 56.8; DB 9; Length 961;
Best Local Similarity 42.88; Pred. No. 0.00043;
Matches 289; Conservative 0; Mismatches 387; Indels 0; Gaps 0;
QY 831 AGTAACCTGGAATATAGGTAATACAAATTCATTATGCGCAATAAGTGTAGGAGGCTAC 890
DB 227 AGCAACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 286
QY 891 GGCTACGTTAGGGGAGCGGTTATTAAAGCCACTACGACTAAGTTAAGCAATGACGCTC 950
DB 287 AACAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 346
QY 951 GGTTATTAACACTTACAAATGCAATTAACAGGTGCTGTTGATTAACCAACAGCGGTGA 1010
DB 347 AGCAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 406
QY 1011 TAATGTAGGTGCTGTAATCTTAAGTGGAGCAATGAGTAACTAACCGGGAATATAGTAA 1070
DB 407 AGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 466
QY 1071 TACAAATTCATTAGCCAGCAATAATATAGGAGCGGTGCTGCTACCTTGGATGAGCGGT 1130
DB 467 AATTAACAACAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 526
QY 1131 TATTAAAGCTACTACGACTAAGTTTAACAGATGATGCTGCTGCTGCTGCTGCTGCTG 1190
DB 527 AACAGCAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 586
QY 1191 TGTAGTGTAGTACCGGCAATAGATATATACCGGTAAATGCCAATAAAGTGTGTTAATCTT 1250
DB 587 AACAGCAGCAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 646
QY 1251 TACCGGAGCAAGTACGGTAAACCGGATAATATAGTAAACAGCGCAGTATTACAGAGGTAAG 1310
DB 647 AGCAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 706
QY 1311 CGTAGGAGCAGGTTTGTCTGCAATATACAGCGGGAGTGTAAAGCGAATGCAATTAACCTT 1370
DB 707 AAGAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 766
QY 1371 AACGGTAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
DB 767 AGCAACAACAATAGCAGCAGCAACAACAACAACAACAACAACAACAACAACAACAACA 826
QY 1431 TGTGTACCAACTCTTCGCAACAGTGAATATAGGAGGAGGATAACATTACAGCGCGGAGG 1490
DB 827 AGCAGCATCAACAGAGCAACAACAACAACAACAACAACAACAACAACAACAACAACA 886
QY 1491 AAGCCTAGCTGCGAAT 1506
DB 887 AACAGCAACAACAAT 902

RESULT 7
US-10-144-771-4707/c
; Sequence 4707, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF

[illegible]

```

RESULT 9
US-10-144-771-33054/c
; Sequence 33054, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144, 771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 33054
; LENGTH: 1118
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-33054

Query Match      2.9%; Score 55.8; DB 9; Length 1118;
Best Local Similarity 42.3%; Pred. No. 0.00079;
Matches 370; Conservative 0; Mismatches 502; Indels 3; Gaps

QY 655 GGAGCAAGTAGCCTTCTCTTGGAGGGCAGTTATTAAAGCCACTACGACTAAGTTAAACA 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 875 GGTCCAGGTACAGAAATTAGGTGCCAGGACGAGGAATAGTGCAGGTGAGGGCAGAGGAGAA 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 715 AATGCGAGCGCGGTATTAAACACTTTACAAATGCAAAATGCAGTATTAAACAGGTGCTGTTGAT 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 815 GATACAGATGAGGAGCAGGTGTAGGTCCAGGACCGAGTGTAGGTGCAGGAGCAGGTGTA 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 775 AACACACAGCGGGTGATGATGTAGGTGCTTTAAATTTAAACGGAGCGTTGAGTCAAGTA 834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 755 GGTGCAGGAGCAGGTGTAGGTGTCAGGAGCAGGAGCAGGTGTAGGTGCAGGAGCAGGAGCA 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 835 ACTGGAATATAGCTAATAACAATTCATTACGACAATAAGTGPAGGAGCAGGTACCGCT 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 695 GGTGTAGCTACAGGAGCAGGTGTAGGTGTAGGTGCAGGACGAGGTGTAGCTACAGGAGCA 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 895 ACGTTTGGGGGAGCGGTTATTAAAGCCACTACGACTAAGTTAACGAATGACGGCTCGGTA 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

QY 616 ACAATAGGTAATACCTCCGACAGTCAATAAGCAATGGAGCAAGTACGCTTTCTCT 675
Db 6347 GCGATAGCAGTACAGCAGTGTATAGCAGTGTAGCAGCAACAGCAGTGTACAGTACGCGATA 6406
QY 676 GGAGGGGCGAGTTATTAAAGCCAGTACGACTAGTTAAACAATGACAGCCGCGGTATTACA 735
Db 6407 GCAGTGCAGCAGCAGCAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAGCAGTGTAG 6466
QY 736 CTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 795
Db 6467 ACAGCAGCAACAGCAGTGTATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 6526
QY 796 GTAGTGTCTTAAATTTAAACCGAGCGTTGAGTCAAGTAACTGGAATATAGGTAATACA 855
Db 6527 GCAACAGCAGTGTATAGCAGCAGCAGCAGTGTATAGCAGTGTATAGCAGTGTATAGCAG 6586
QY 856 AATTCATTAGCAGCAATAGTGTAGCAGCAGTGTAGCAGCAGTGTAGCAGCAGCAGCAGT 915
Db 6587 GCGACAGCAGCAGTGTATAGCAGCAGTGTATAGCAGCAGCAGTGTATAGCAGCAGCAGCAG 6646
QY 916 AAAGCCACTAGCAGTAACTTAAACGAGTGTAGCAGCAGTGTATAGCAGCAGCAGTGTAT 975
Db 6647 GCAGTGTATAGCAGCAGCAGCAGTGTATAGCAGCAGCAGTGTATAGCAGCAGCAGCAGT 6706
QY 976 TTAACAGTGTCTGTATTAACACACAGCAGCAGTGTATAGCAGCAGCAGTGTATAGT 1035
Db 6707 GAAGTGACAGTGTATTAATAGTGTAGCAGCAGCAGTGTATAGCAGCAGCAGCAGTGT 6766
QY 1036 GGAGCATTGAGTCAAGTAAACCGGGAATATAGGTAATACAAATTTAGCAGCAGCAGTAA 1095
Db 6767 ACAGCAGTGTATAGTGTAGCAGCAGTGTATAGCAGCAGCAGTGTATAGCAGCAGCAGT 6826
QY 1096 ATAGCAGCAGTGTGTGTACCTTGGATGGAGCGGTTATTAAAGCTACTACGACTAAAGTT 1155
Db 6827 GTGATAGCAGCAACAGCAGTGTATAGTGTAGCAGCAGTGTATAGCAGCAGCAGCAGT 6886
QY 1156 ACAGATGATGCGTCAAGTATTAATTTAGCAATCCTGTAGTGTATAGTGTATAGTGTAT 1209
Db 6887 GTGATAGCAGCAACAGCAGTGTATAGTGTAGCAGCAGTGTATAGTGTATAGTGTATAG 6946
QY 1210 ATAGATATACCGGTAATGCGAATAAAGGTGTGTGTATCTTTACCGGAGCAAGTACGTA 1269
Db 6947 GCAGTGTAGCAGCAACAGCAGTGTATAGCAGCAGTGTATAGCAGCAGTGTATAGTGTAT 7006
QY 1270 ACCGATATATAGTGTACACCGCAGTGTATAGCAGCAGTGTATAGCAGCAGTGTATAG 1329
Db 7007 ACAGCAGCAATAGCGGTGTAGCAGCAGCAGCAGTGTATAGCAGCAGTGTATAGCAGCAG 7066
QY 1330 CAATACAAAGCGGAGTAGTAAAGCGAATGCAATTAACCTTAAACGGGTAATGCGTCACT 1389
Db 7067 ACAGCAGTGTAGCAGCAACAGCAGCAGTGTATAGCAGCAGTGTATAGCAGCAGTGTAT 7126
QY 1390 GT---ACATTTACCGGTGTATAGTGTATAGCAGTGTATAGTGTATAGTGTATAGTGTAT 1446
Db 7127 GTGACAGCAGTGTATAGCAGCAACAGCAGTGTATAGCAGTGTATAGCAGTGTATAGCAG 7186
QY 1447 GCAACAGTGTATAGCAGCAGCAGTGTATAGCAGTGTATAGCAGTGTATAGCAGTGTAT 1486
Db 7187 GCAGTAAATAGTGTAGCAGCAGCAGCAGCAGTGTATAGCAGTGTATAGCAGTGTATAG 7226

```

RESULT 15

```

US-10-363-798-1/c
: Sequence 1, Application US/10363798
: GENERAL INFORMATION:
: APPLICANT: Kong, Xiangyin
: APPLICANT: Xiao, Shangxi
: APPLICANT: Zhao, Guoping
: APPLICANT: Yu, Chuan
: APPLICANT: Hu, Lian
: TITLE OF INVENTION: METHOD OF DIAGNOSING AND TREATING DENTINOGENESIS IMPERFECTA
: TYPE II USING DENTIN SIALOPHOSPHOPROTEIN GENE AND CODED

```

```

: TITLE OF INVENTION: PRODUCT THEREOF
: FILE REFERENCE: 9548.78USWO
: CURRENT APPLICATION NUMBER: US/10/363,798
: CURRENT FILING DATE: 2003-03-05
: PRIOR APPLICATION NUMBER: CN 00125042.6
: PRIOR FILING DATE: 2000-09-05
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 8201
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-363-798-1

```

```

Query Match      2.8%   Score 53.2; DB 8; Length 8201;
Best Local Similarity 41.1%; Pred. No. 0.0053;
Matches 379; Conservative 0; Mismatches 543; Indels 0; Gaps 0;

QY 102 TGGAGTCTCTTGGTGTGCAAGAACCGTAACGCTGATGGTGACAGAGCTTGCAGCCGGAAC 161
Db 7459 TGCTGTCACTGCTGCTGCTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7400
QY 162 AAATATAGTCTGAGCGCGGTGCTTTGTAGCGGGTCTTACTTTACAATATACCGGTGC 221
Db 7399 CACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7340
QY 222 TTTTACGGTTACTGATGCTGACGTAAGTGTTCGTCGATTAAGTTTAAATTTTGCAGC 281
Db 7339 TGTCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7280
QY 282 AGGCTCTTTTTCAGTAATCTGCTGATATTTTCAATAGTTTCAGTGTAGTAGTACGGGAGGAGC 341
Db 7279 TGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7220
QY 342 TAATAACTTTGCGATTAATATTGATGATGGTTTAAACCTTAACTTTACAGAGTACCGGTAC 401
Db 7219 CACTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7160
QY 402 TGCAGCCTCAGGTTGCAAAATCCTGCTGTTGTTATTCCAAAGTGGACAAAGCTGCTGCTATAA 461
Db 7159 TATCACTGCTGTTGCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7100
QY 462 TACATATAGTCTTTAGGTAATTAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
Db 7099 TGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7040
QY 522 TTCAGATCCAGATGATATTAGGACCAATAACGCTTTCAGGAGAAATATAGATGGAGGAGTAT 581
Db 7039 TGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6980
QY 582 AATACTGCAATATACAGATGCTGCCATTAAACGGAACAATAAGTAACTAATCCGGCAGC 641
Db 6979 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6920
QY 642 TCAAAATAGCATTTGGAGCAAGTACGCTTCTCTTTGAGGGGAGGATTTAAAGCCACTAC 701
Db 6919 TTTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6860
QY 702 GACTAAGTTAAACAAATGACAGCGCGGTATTAAACACTTACAAATGCAAAATGCAAGTATTAA 761
Db 6859 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6800
QY 762 AGGTGCTGTTGATTAACACACAGCGGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 821
Db 6799 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6740
QY 822 GTTGAGTCAAGTAACTGGAAATATAGTAACTAATTTATAGCCGCAATAGTGTAGT 881
Db 6739 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6680
QY 882 AGCAGTACGGCTACGTTAGGGGAGCGGTTATTAAAGCCACTTACGACTAAAGTTAAACGAA 941
Db 6679 CACTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6620

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 17, 2003, 09:22:14 ; Search time 5424 Seconds
(without alignments)
5732.916 Million cell updates/sec

Title: US-09-800-065-1

Perfect score: 1920

Sequence: 1 atggcgaaatttctctaaa.....cgttggagctaaacaatag 1920

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.6	3.6	805	17 CNS04RW2	AL304427 Tetraodon
2	63.6	3.3	443	17 FR0008252	Z92062 F.rubripes
3	63.4	3.3	711	13 BM588321	BM588321 170006873
4	62.4	3.2	644	13 BM605179	BM605179 170006870
5	61	3.2	1101	17 CNS0039G	AL063921 Drosophila
6	60.4	3.1	641	13 BM181884	BM181884 fv51b11.y

7	60.4	3.1	935	17 CNS033D4	AL225985 Tetraodon
8	60	3.1	536	17 FR0036552	AL124061 Fugu rubr
9	59.6	3.1	454	17 AF046361	AF046361 AF046361
10	58.2	3.1	819	10 AW128683	AW128683 fe36a05.y
11	58.4	3.0	699	13 BM630395	BM630395 170006875
12	57.8	3.0	785	13 BI694638	BI694638 603347867
13	57.8	3.0	857	12 BG837033	BG837033 Zm08_10h0
14	56.4	2.9	1076	17 CNS05HXN	AL338180 Tetraodon
15	56.2	2.9	787	13 BJ282421	BJ282421 BJ282421
16	56.2	2.9	970	17 CNS03H6V	AL243904 Tetraodon
17	56.2	2.9	978	17 CNS040R1	AL269254 Tetraodon
18	55.4	2.9	621	17 BH886515	BH886515 LB00838a.
19	55.4	2.9	607	17 BH887503	BH887503 LB01480a.
20	55	2.9	599	13 BM601700	BM601700 170006870
21	54.8	2.9	776	12 BG750043	BG750043 602708813
22	54.6	2.8	526	17 CNS03CM0	AL237969 Tetraodon
23	54.6	2.8	1003	17 CNS04QUZ	AL303092 Tetraodon
24	54.6	2.8	1175	13 BI872945	BI872945 603398115
25	54.4	2.8	571	17 AQ919110	AQ919110 RPCI-23-2
26	53.6	2.8	747	13 BG918691	BG918691 602818643
27	53	2.8	1038	17 CNS0340G	AL226825 Tetraodon
28	52.8	2.8	526	17 AZ312601	AZ312601 LM0028104
29	52.8	2.8	619	17 FR0006944	Z90754 F.rubripes
30	52.8	2.8	946	9 AI069309	AI069309 mgae00006c
31	52.6	2.7	980	17 CNS01PHD	AL154817 Anopheles
32	52.4	2.7	494	10 AW119988	AW119988 707006605
33	52.4	2.7	527	10 AV962678	AV962678 AV962678
34	52.4	2.7	645	13 BJ404348	BJ404348 BJ404348
35	52.2	2.7	551	17 AZ613984	AZ613984 LM0442003
36	52	2.7	378	13 BJ370516	BJ370516 BJ370516
37	52	2.7	479	17 BH211305	BH211305 Sml-50G22
38	51.8	2.7	437	17 AQ246813	AQ246813 HS_3009.A
39	51.6	2.7	906	17 CNS02ILL	AL199074 Tetraodon
40	51.4	2.7	606	13 BJ314050	BJ314050 BJ314050
41	51.4	2.7	726	14 BM944226	BM944226 UI-M-EHOP
42	51.4	2.7	786	13 BI106945	BI106945 602894749
43	51.2	2.7	740	10 AV973760	AV973760 AV973760
44	51	2.7	723	17 BH182027	BH182027 O20_K.19-
45	51	2.7	723	17 CNS07NIN	AL618977 T3 end of

ALIGNMENTS

RESULT 1	CNS04RW2	805 bp	DNA	linear	GSS 24-MAY-2000
CNS04RW2	Tetraodon nigroviridis genome survey sequence PUC-ORI end of clone 005C20 of library H from Tetraodon nigroviridis, genomic survey sequence.				
LOCUS	AL304427				
DEFINITION	AL304427.1 GI:8192034				
ACCESSION	GSS: genome survey sequence.				
VERSION	Tetraodon nigroviridis.				
KEYWORDS	Tetraodon nigroviridis				
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.				
ORGANISM	1 (bases 1 to 805)				
REFERENCE	Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.				
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 805)				
REFERENCE	Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.				
AUTHORS	Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis				
TITLE	Unpublished				
JOURNAL					


```

/db_xref="taxon:7165"
/clone="19600449697628"
/clone.lib="A.Gam.ad.cdna.blood1"
/dev_stage="Adult"
/lab_host="DH10b"
/Note="Vector: pSport1; Site 1: Sall; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT 305 a 234 c 135 g 37 t
ORIGIN
Query Match 3.3%; Score 63.4; DB 13; Length 711;
Best Local Similarity 45.5%; Pred. No. 2.9e-06;
Matches 315; Conservative 0; Mismatches 366; Indels 12; Gaps 2;
QY 95 CCGGTAGTGCAGTCTTGGTGTGCAAGAACCGTAACCTGCTGATGTGCAGAGCTTGCGAG 154
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
685 CTGTTGCTGCAGATGTTGTTGTTGCTGCGCGGCTGCTGCTGCGCGGCTGCTG 626
QY 155 CCGGACAAATATAGTCTCGAGCGGTGCTTTTGTAGCGGTGCTACTTTACAAATA 214
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
625 CTGCTGCAGTTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 566
QY 215 CCGGTGCTTTTACGGTTACTGATGCTGAGTAAGTCTTTCGTCATAGATTAATAAT 274
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
565 CTGCTGTTGCTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 506
QY 275 TTGCAGCAGGCTTTTTCAGTAACCTGCTGATGATTTCAATAGGTCAGTGTAGATCG 334
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 446
QY 335 GAGGAGCTAAATAACTTGACGATTAATATGATGATGTTTAACTTAACCTTAACAG 394
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 TTGCTGC-----TGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 395
QY 395 CCGGTACTGCAGCTACGGTGCAGAAATCCCTGCTGTTGTTATTCAGGTGGACAGCTG 454
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 CTGCTGTTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 335
QY 455 CTAAATAACATATACCTGTTTGTAGTAAATATACCTAGTGTGAGCGAATGCCGTTTGA 514
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
334 CTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 275
QY 515 CTATTGCTTCAGATCCAGATGATTAGGACCAATACCGTTTCAGGAATATAGATGGAG 574
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 TAATTGTTGCAATTT---GATGCTGTTCATGCTGCTGCTGCTGCTGCTGCTGCTG 218
QY 575 GAGGTATAATACTGACATACAGATGCTGCCATTAAACGGAACAATAGGTAATACTAATC 634
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 ATGTTGTGATGCTGCACCTGATGCTGCTGTTTCTGAGCAAGTGCATTTCCTGATG 158
QY 635 CCGGAGCTCAATAAGCATTTGAGCAAGTACGCTTCTCTTGGAGGGGAGTATTAAAG 694
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 CATCTGCTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 98
QY 695 CCACACGACTAAGTTAACAATAGCAGCGCGGTATTAACTTACAATGCAAAATGACAG 754
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 CTGTTGCTGCTGTTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 38
QY 755 TATTAAACAGGCTGCTGTGATAACACACAGCGG 787
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
37 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5
RESULT 4
BM605179/c 644 bp mRNA linear EST 25-FEB-2002
LOCUS BM605179
DEFINITION 17000687075984 A.Gam.ad.cdna.blood1 Anopheles gambiae cDNA clone
19600449700619 5', mRNA sequence.
ACCESSION BM605179

```

```

VERSION BM605179.1 GI:18903283
KEYWORDS EST
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 644)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AY row: C column: 05
Seq primer: M13 Reverse
FEATURES
location/Qualifiers
source
1..644
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449700619"
/clone.lib="A.Gam.ad.cdna.blood1"
/dev_stage="Adult"
/lab_host="DH10b"
/Note="Vector: pSport1; Site 1: Sall; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT 284 a 202 c 119 g 39 t
ORIGIN
Query Match 3.2%; Score 62.4; DB 13; Length 644;
Best Local Similarity 46.4%; Pred. No. 5.2e-06;
Matches 204; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
QY 97 GGTAGTGGAGTCTCTGGTCTGCAAGACCGTAACCTGCTGATGTGCAGAGCTTGCGAGC 156
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 GTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 394
QY 157 GGAACAAATATAGTCTCGAGCCGGTCTTTTGTAGCGGGTCTACTTTACAATATACC 216
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 GCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 334
QY 217 GGTGCTTTTACGGTTACTGATGCTGACGTAAGTGTTCGTCATTAGATTAATAATTTT 276
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 CTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 274
QY 277 GCAGCAGGCTCTTTTTCAGTAACCTGCTGATATTTCAATAGGTTCACTGCTAGATACG 336
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
273 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
QY 337 GGAGCTAATAAATTCGAGTTAATATTTGATGTTTAACTTAACTTAACTTAACTTAACT 396
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 GTTGTGTTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154
QY 397 GGTACTGCAGCTACGGTGCAGAAATCTCGGTTGTTATTCAGAGTGGACAACTGCTG 456
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 GCTGTGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 94
QY 457 AATAATACATATACCTTTAGGTAATAATACTAGTGGAGCGGAATCCGCTTTGACT 516
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
93 GATGCTCTGCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 34
QY 517 ATTGCTTCAGATCCAGATCT 536
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 33 GTTGATTTCTCGGACATTT 14

RESULT 5
CNS00396/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
AL063921
VERSION
AL063921.1 GI:4941778
SOURCE
GSS.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/TheBDGP/Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES
source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPI-98"
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

Query Match 3.2%; Score 61; DB 17; Length 1101;
Best Local Similarity 16.9%; Pred. No. 1.7e-05;
Matches 119; Conservative 290; Mismatches 293; Indels 3; Gaps 1;

Qy 1123 GGACGGTTTAAAGTACTACGAGTAAGTTAACAGATGCGTCAGTATTAATTT 1182
|||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1101 GKARRGGDTWDRTRKDDWTKWTKWTKDRADRRWAGDADRWAGDGTWTTATW 1042
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1183 ACGAATCTGTAGTAGTAACCGGAGCAATAGATATACCGTATGCCATGAAGGTGTG 1242
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1041 WWWWWATWDTWDDKWWWWATAAKTDTAFTWRTAWRADRWAGDRGAKRDRDAATDAG 982
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1243 GTAATCTTTACCGGAGCAAGTACCGTAACCGTAATAGTATACCGCGAGTATTAGCA 1302
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 981 AGRDGGKRRKDKDRKDDGDKGKKKKAAKAATWTKWDDWDDKDKWDDGAKDRK 922
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1303 GAGTAAGCTAGACGAGGTTTGTGCAATACAGCGGAGTAGTAAAGCGAATGCA 1362
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 ADDDAGDKDDGDKGADDDTDTGTDKDDDDKDKWDDWDRKAGTWGATWAAATDWWW 862
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1363 ATAACTTAACGGATAGCGTCAGTAGTAACATTTACCGGTGATAGTAGCGTACAGGT 1422
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 861 GNADADWTTWDAADWADWDDAWKWKDDAWANGARTADRRDDWDRAGKRGGAARR 802
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1423 AGTATAGGTGTACAGAACTCTTCGCAACAGTGAATATAGGAGCAAGTAACATTACGA 1482

```

```

Db 801 DKRRADDRDAADRRDDAATTTTTRTDDWKWKTDTTWTRWAADRTWDRDDDDRD 742
Qy 1483 GCCGGAGGAGCCCTAGCTGCGAATAATAGATTTTGGAGCTGCGAGTAATTTAGAGTTT 1542
Db 741 RAGTAGRKWRTWKR---RWKRTRWDDADDDTARDRRRRGGDDGADGAKGKTKGRK 685
Qy 1543 AACGGTCTGCGGTAGAAATTAACCTTAATCGAACTATATGCAACGCGTAAATAGCT 1602
Db 684 RRRDRATWDRTDAAWADAANWTTTDTDDWDRKRRRGARRRRRTTARAADWWTWKA 625
Qy 1603 ACACCTTAATATTAATGCTGCTGCTACAGTGATTCGAATGATGTAGTATAGGACAGTT 1662
Db 624 WDWAQWQKTRADRWDAADTTWTDARKADRWAKARAWARRDRARAARDRRWTTNGK 565
Qy 1663 GCACAAATTAACATTCAAAATAAAGATTTTGTCTAATAATGCTAAGAACGCTGATGTT 1722
Db 564 TTTATTTWAARAANWAWWATTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTW 505
Qy 1723 GATATATTAGCGCTCAGGCGATTAGTTTAAAGAGAGAGCTTCACGCTTTTCTTAGCT 1782
Db 504 TATWAATAAWAAAAAATAATTTTTTTWTWTWTWTWTWTWTWTWTWTWTWTW 445
Qy 1783 AACCTTAGCTACAGATGATAGCTTATCTACTCTTAAATCAAT 1827
Db 444 ATTTTWTWTWTWAAATATTTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 400

RESULT 6
BM181884/c
LOCUS
DEFINITION
rv51b11.y1 Sugano SJD adult male Danio rerio cDNA clone 5412044 5'
similar to contains element TAR1 repetitive element ;, mRNA
sequence.
ACCESSION
BM181884
VERSION
BM181884.1 GI:17512842
KEYWORDS
EST.
SOURCE
zebrafish.
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 641)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from AmerSham
High quality sequence stop: 527.
Location/Qualifiers
1. 641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SJD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"

```



```

SOURCE      Takifugu rubripes.
ORGANISM    Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1 (bases 1 to 536)
AUTHORS    Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrani,Y., Williams,G. and Brenner,S.
TITLE      Direct Submission
JOURNAL    Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hmp.mrc.ac.uk
COMMENT    Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESC:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES    Location/Qualifiers
            source
            1..536
            /organism="Takifugu rubripes"
            /db_xref="taxon:31033"
            /clone="032K14aF12"
            /clone_lib="cosmid 032K14"
BASE COUNT 44 a 136 c 165 g 157 t 34 others
ORIGIN
Query Match 3.1%; Score 60; DB 17; Length 536;
Best Local Similarity 47.3%; Pred. No. 2.2e-05;
Matches 192; Conservative 0; Mismatches 209; Indels 5; Gaps 1;
Qy 110 TTGCTGCTGCAAGAACGTAAGTCTGCTGATGCTGAGAGCTTGCAGCGGACAAATATAG 169
Db 41 TAGCTGCTGCTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
Qy 170 GTCCTGAGCGGCGTCTTTGTAGCGGCTCTACTTTACAAATATACCGGTCGCTTTACGG 229
Db 101 CTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 160
Qy 230 TTACTGATGCTGAGCTAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 289
Db 161 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 220
Qy 290 TTTGAGTAAGTGTGATTTTCATTAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 349
Db 221 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280
Qy 350 TTGCAATTAATATGATGATGCTTTACCTTAACCTTAACCTTAACCTTAACCTTAAC 404
Db 281 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 340
Qy 405 AGCTACGGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 464
Db 341 TGCTGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
Qy 465 ATATAGCTTTTAAAGTAAATTAACCTTAAGTGTGAGCGGAATCCCGGT 510
Db 401 TGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446

RESULT 9
AF046361/c AF046361 Mus musculus 129Sv/Ev Mus musculus genomic clone OSTI5546,
LOCUS      DNA sequence.
DEFINITION
ACCESSION  AF046361
VERSION     AF046361.1 GI:3005232
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE   1 (bases 1 to 454)
AUTHORS    Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L.,
            Person,C. and Sands,A.T.
TITLE      Disruption and sequence identification of 2,000 genes in mouse
            embryonic stem cells
JOURNAL    Nature 392 (6676), 608-611 (1998)
MEDLINE    98219085
COMMENT    Contact: Zambrowicz BP
            OmniBank
            Lexicon Genetics Incorporated
            4000 Research Forest Drive, The Woodlands, TX 77381, USA
            Email: matrisa@lexgen.com
            OmniBank Sequence Tag
            Class: exon-trapped.
FEATURES    Location/Qualifiers
            source
            1..454
            /organism="Mus musculus"
            /strain="129Sv/Ev"
            /db_xref="taxon:10090"
            /clone="OSTI5546"
            /clone_lib="Mus musculus 129Sv/Ev"
            /cell_type="embryonic stem cell"
BASE COUNT 155 a 104 c 94 g 70 t 31 others
ORIGIN
Query Match 3.1%; Score 59.6; DB 17; Length 454;
Best Local Similarity 51.7%; Pred. No. 2.6e-05;
Matches 138; Conservative 0; Mismatches 123; Indels 6; Gaps 1;
Qy 189 TGTAGCGGTTCTACTTACAAATATACCGTGTCTTTACGGTTACTGCTGACGTAAAG 248
Db 295 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 236
Qy 249 TGTTCGTGCATTAGATTAAATAATTTGCAGCAGGCTTTTTTTCAGTAACTGCTGATAT 308
Db 235 TGNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 176
Qy 309 TTCATTAGTTCAGTGTGATAGTACGGAGGAGCTTAATAAACTTGCAGTTAATATTGATGA 368
Db 175 TTNTGTTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
Qy 369 TGGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 428
Db 121 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62
Qy 429 GTTATTCGAAGTGGACAAGCTGCTGC 455
Db 61 NNCAGNCAAGATAAAGNACCGCTCC 35

RESULT 10
AW128683/c AW128683
LOCUS      819 bp mRNA linear EST 07-JUN-2001
DEFINITION fa36a05.y1 zebrafish Washu MPIMG EST Danio rerio cDNA clone
IMAGE:3740912.5 similar to TR:Q9Z0W6 Q9Z0W6 PAX TRANSCRIPTION
ACTIVATION DOMAIN INTERACTING PROTEIN PTIP ; contains element TARI
repetitive element ;, mRNA sequence.
ACCESSION  AW128683
VERSION     AW128683.1 GI:6116587
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostarlophysi; Cypriniformes
            ; Cyprinidae; Danio.
            1 (bases 1 to 819)
REFERENCE   1 (bases 1 to 819)
AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
            WashU Zebrafish EST Project 1998
TITLE      WashU Zebrafish EST Project 1998

```

JOURNAL
COMMENT

Unpublished (1998)
Other_ESRs: fe36a05.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
www.rzdp.de)

Seq primer: T3 ET from Amersham
High quality sequence stop: 485

[illegible]

```

high quality sequences kept: 483.
Location/Qualifiers
i. .819
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="IMAGE:3740912"
/clone_lib="Zebrafish WashU MPMG EST"
/sex="mixed"
/tissue.type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACATGCTTCAATCCGAGGCGCCCTTTTTTTTTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones for
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality

```

	control."		
BASE COUNT	307 a	253 c	
ORIGIN			
		142 g	111 t
			6 others

Query Match	3.1%	Score 59.2;	DB 10;	Length 819;
Best Local Similarity	47.5%;	Pred. No. 4.6e-05;		
Matches 209; Conservative	0;	Mismatches 228;	Indels 3;	Gaps 1;

[illegible]

Qy	397	GGTACTGAGCCTACGCTGCAAAATCCCTGCTGTTATTATCCAAAGTGGACAAGCTGCTGCT	456
Db	292	CATATTGCTGCTGATGCTGCAAGGCGAGCCCTCGGAAGCTTTGGGGTCGCAATTTGTTGTG	233
Qy	457	AATAATACATATACTGCTTTAGGTAAATATAACTCTAGGTGAGCGAATGCCGGTTTGACT	516
Db	232	GGAATGATGTTGTTGCTGCTGCTGTTGTTGCTGCTGCTGCTGCTGTTGATGTGTGGAAAT	173
Qy	517	ATTGCTTCAGATCCAGATGT	536
Db	172	ACATCTGCTGATGCATCTGT	153
RESULT 11			
BM630395/c			
LOCUS			
DEFINITION	17000687501370 A.Gam.ad.cDNA1 Anopheles gambiae EST 26-FEB-19600449647565 5', mRNA sequence.		
ACCESSION	BM630395		
VERSION	BM630395.1		
KEYWORDS	GI:18929906		
SOURCE	EST.		
ORGANISM	African malaria mosquito.		
REFERENCE	Anopheles gambiae		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.		
TITLE	1 (bases 1 to 699)		
COMMENT	Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Chao,R., Collins,F.H., Venter,J.C. and Hoffman,S.L. Celera Anopheles gambiae EST project Unpublished (2002) Contact: Holt R.A. Celera Genomics 45 W. Gude Dr., Rockville, MD 20850, USA Tel: 2404533151 Fax: 2404534580 Email: HoltRA@celera.com Plate: NU010049WA row: A column: 03 Seq primer: M13 Reverse. Location/Qualifiers 1..699		

FEATURES
SOURCE

```

1. 1023)
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449647565"
/clone_lib="A.Gam.ad.cdna1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cdna inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
261 a 230 c 155 g 53 t
BASE COUNT
ORIGIN

```

Query Match	3.0%	Score 58.4;	DB 13;	Length 699;
Best Local Similarity	49.7%	Pred. No. 7e-05;		
Matches 149;	Conservative	0;	Mismatches 151;	Indels 0;
				Gaps 0;

Qy	125	CCGTAAC	TGCTG	TGATG	GTGCAG	AGCTTTG	CAGCCGGA	CAAAAT	ATAGCTC	TCTGAGCGCGTG	184
Db	421	CTGTTG	TGTTG	TGCTG	TGTTG	TGCTG	CTGCTG	CTGCTG	TGTTG	TGCTG	TGTTG
Qy	185	CTTTTGT	AGCGG	TTTCA	CAATAT	ACCGG	TCTTTT	ACGGTT	TACTG	TCTGACG	244
Db	361	CTGTTG	TGTTG	TGCTG	TGTTG	TGCTG	CTGCTG	TGCTG	TGTTG	TGCTG	TGTTG
Qy	245	TAAGTGT	TCTGCA	TATAG	ATTTAA	ATAATTT	TTCGAC	AGGTCT	TTTTTTC	AGTAAC	TCGGTG
Db	301	TAGCTGT	TGCTG	TGCTG	TGCTG	TGCTG	TGCTG	TGCTG	TGTTG	TGCTG	TGCTG

QY	342	TATAAACTTGCAGTTTAATTATGTGATG 370 Db 122 TGCTGCTGTAGCTGTTTGCTGCTGCTG 94
RESULT 13		
BG837033/c		
LOCUS		
DEFINITION	BG837033 Zm08_10h09_A 857 bp mRNA linear EST 25-MAY-2001	
ACCESSION	Zm08_AAFc_ECORC_Fusarium_graminearum_inoculated_corn_ear_Zea mays	
VERSION	CDNA clone Zm08_10h09, mRNA sequence.	
KEYWORDS	BG837033 1 GI:14203356	
SOURCE	EST.	
ORGANISM	Zea mays. Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 857) Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott .D. and Tinker,N.A. Expressed Sequence Tags from Developing Maize Kernels Six Days after Silk Channel Inoculation with Fusarium graminearum Unpublished (2001) Contact: Harris, Linda J. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6, CANADA Tel: (613) 759-1314 Fax: (613) 759-6566 Email: harrislj@em.agr.ca. Location/Qualifiers 1..857 /organism="Zea mays" /cultivar="Co430" /db_xref="taxon:4577" /clone="Zm08_10h09" /clone_lib="Zm08_AAFc_ECORC_Fusarium_graminearum_inoculate d_corn_ear" /tissue_type="Developing kernels (sibcrossed)" /dev_stage="10-11 days post-silk emergence" /notes=Vector: Bluescript SK+/XhoI-EcoRI; Site_1: EcoRI; Site_2: XhoI; Field-grown maize ears were silk channel-inoculated in the morning (~10 am) with 1 ml of a Fusarium graminearum macroconidial suspension (500,000 spores/ml) and whole ears were collected and immediately frozen in liquid nitrogen 6 days later." BASE COUNT 273 a 220 c 217 g 146 t 1 others ORIGIN	
Query Match	3.0%; Score 57.8; DB 12; Length 857;	
Best Local Similarity	48.3%; Pred. No. 0.00011;	
Matches 161; Conservative	0; Mismatches 172; Indels 0; Gaps 0;	
QY	169	GGTCCTGAGCGCGTGCTTTTGTAGCGGGTCTACTTTACAATATACGGGTGCTTTTACG 228 Db 443 GTTGCTGCTGCTGTCGATTTGCTGCATCTGTTGCTGTGAGCGTGCATTTGCTGCTGCT 384
QY	229	GTTACTGATGCTGACGTAAGTGTTCGTCATTAGATTTAAATATTTGCAGCAGGCTTT 288 Db 383 GCTGCTGCATTTGCTGCATCTGTTGCTGCTGAGCGTGCATTTGCTGCTGCTGCTGCA 324
QY	289	TTTTTCAGTAACTGGTGATATTTTCATTAGGTTTCAGTGTGATACGCGGAGGAGCTAATAA 348 Db 323 TTGCTGCATCTGTTGCTGCTGAGCGTGCATCTGTTGCTGCTGCTGCTGCTGCTATT 264
QY	349	CTTGCAGTTAATATGATGATGGTTTAACTTTAACTTTAACAGGTACCGGTACTGCGACCC 408 Db 263 GCTCACTGTTGTTGTTGATGCTGCATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCA 204

[illegible]

Search completed: April 17, 2003, 13:33:25
Job time : 5439 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:00:54 ; Search time 40 seconds
(without alignments)
2062.053 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKLQKAIQKGLKTA.....MIELSLKIIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/cgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/cgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/cgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3022	100.0	619	AAE10890	Rickettsia felis o
2	333.5	11.0	1026	AAE10890	Caulobacter S-laye
3	331	11.0	1026	AAW37490	Caulobacter cresce
4	331	11.0	1026	AAW44757	Caulobacter cresce
5	323.5	10.7	1026	AAW48993	rsaa S-lyae prote
6	322	10.7	1180	AAW01845	Haemophilus influe
7	322	10.7	1188	AAW01844	Haemophilus influe
8	316.5	10.5	1222	AAW01830	H. influenzae stra
9	316.5	10.5	1228	AAW01828	Haemophilus influe
10	310	10.3	1095	AAW01835	Haemophilus influe

11	310	10.3	1101	21	AAW01834	Haemophilus influe
12	303	10.0	1267	11	AAW04232	Rickettsia rickett
13	299.5	9.9	1338	14	AAW41731	High molecular wei
14	293.5	9.9	1598	18	AAW30291	Non-typeable Haemo
15	299.5	9.9	2732	22	AAW52855	Escherichia coli p
16	298.5	9.9	1529	14	AAW41732	High molecular wei
17	298.5	9.9	1601	18	AAW30292	Non-typeable Haemo
18	293.5	9.7	969	21	AAW01827	Haemophilus influe
19	293.5	9.7	975	21	AAW01826	Haemophilus influe
20	292	9.7	1612	19	AAW65088	R. prowazekii S-la
21	290.5	9.6	1221	21	AAW01825	Haemophilus influe
22	290.5	9.6	1227	21	AAW01824	Haemophilus influe
23	286	9.5	1477	14	AAW41724	High molecular wei
24	280	9.3	1036	21	AAW01849	Haemophilus influe
25	280	9.3	1477	14	AAW41728	High molecular wei
26	280	9.3	1477	21	AAW01848	Haemophilus influe
27	279	9.2	1477	18	AAW30294	Non-typeable Haemo
28	278	9.2	1477	15	AAW63506	Haemophilus high m
29	276	9.1	871	21	AAW95550	Chlamydia pneumoni
30	276	9.1	1407	23	AAW90541	Chlamydia pneumoni
31	273.5	9.1	1095	21	AAW01847	Haemophilus influe
32	273.5	9.1	1536	14	AAW41723	High molecular wei
33	273.5	9.1	1536	14	AAW41725	High molecular wei
34	273.5	9.1	1536	15	AAW63505	Haemophilus high m
35	273.5	9.1	1536	21	AAW01846	Haemophilus influe
36	273	9.0	1749	22	AAW72314	Glutamic acid rich
37	272.5	9.0	1536	18	AAW30293	Non-typeable Haemo
38	269	8.9	1005	21	AAW01833	Haemophilus influe
39	269	8.9	1011	21	AAW01832	Haemophilus influe
40	263	8.7	1648	23	AAW54925	Lactococcus lactis
41	262	8.7	992	21	AAW01843	Haemophilus influe
42	262	8.7	998	21	AAW01842	Haemophilus influe
43	260	8.6	943	19	AAW81745	M. tuberculosis im
44	260	8.6	943	19	AAW64378	Mycobacterium tube
45	260	8.6	943	20	AAW39175	M. tuberculosis an

ALIGNMENTS

RESULT 1

AAE10890

ID AAE10890 standard; Protein; 619 AA.

XX

AC AAE10890;

XX

DF 18-DEC-2001 (first entry)

XX

DE Rickettsia felis outer membrane protein (rompA).

XX

KW Outer membrane protein; rompA; antibacterial.

XX

OS Rickettsia felis.

XX

PN WO200166691-A2.

XX

PD 13-SEP-2001.

XX

PF 06-MAR-2001; 2001WO-US07820.

XX

PR 06-MAR-2000; 2000US-187323P.

XX

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Bouyer DH, Crocquet-Valdes P, Stenos J, Walker DH;

XX

DR WPI; 2001-596829/67.

XX

DR N-PSDB; AAD18232.

XX

PT Novel isolated Rickettsia felis outer membrane protein and polynucleotide for modulating expression of the protein in a host cell

XX

appicant

PS Claim 27; Page 62-64; 73pp; English.

CC The present invention relates to an isolated Rickettsia felis outer
 CC membrane protein (rompA) and its polynucleotide. The rompA gene is
 CC useful for modulating expression of the protein in a host cell. The
 CC rompA gene is useful as probe or for the design of primers to obtain
 CC DNA encoding the protein by either cloning and colony/plaque
 CC hybridisation or amplification using PCR. The present sequence is
 CC Rickettsia felis outer membrane protein (rompA).

XX Sequence 619 AA;

Query Match 100.0%; Score 3022; DB 22; Length 619;
 Best Local Similarity 100.0%; Pred. No. 2.9e-186;
 Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANISKLKFORAIQKGLKLTFTTSTAAILMTGSGVLTADGAELAGTNIQGP 60
 Db 1 MANISKLKFORAIQKGLKLTFTTSTAAILMTGSGVLTADGAELAGTNIQGP 60

Qy 61 GAFVAGSTLQVTGFTVTDADVSRALDNNFAAGLFSVTGDISLGSVVDGKANKLAVN 120
 Db 61 GAFVAGSTLQVTGFTVTDADVSRALDNNFAAGLFSVTGDISLGSVVDGKANKLAVN 120

Qy 121 IDDLGLTLTGTGTAAYGANPALLFOGGQAAANNNTYALGNITLGGANAGLTIASDPDVL 180
 Db 121 IDDLGLTLTGTGTAAYGANPALLFOGGQAAANNNTYALGNITLGGANAGLTIASDPDVL 180

Qy 181 GPITLAGNIDGGIITDNTDAIINGTNTGNTNPAQAISIGASTLSLGGAVIKATTTKLTNA 240
 Db 181 GPITLAGNIDGGIITDNTDAIINGTNTGNTNPAQAISIGASTLSLGGAVIKATTTKLTNA 240

Qy 241 APVLTITNANAVLTGAVDNTTGGDDVGVNLNGALQSQTGNIGNTSLATISVGAGTATL 300
 Db 241 APVLTITNANAVLTGAVDNTTGGDDVGVNLNGALQSQTGNIGNTSLATISVGAGTATL 300

Qy 301 GGAVIKATTTKLTNAASVLTITNNAVLTGAVDNTTGGDVGNNVNLGGALSQVTGNIGNTNS 360
 Db 301 GGAVIKATTTKLTNAASVLTITNNAVLTGAVDNTTGGDVGNNVNLGGALSQVTGNIGNTNS 360

Qy 361 LATINIGAGVATLGGAVIKATTTKLTDDASVLIFTPVVTGAIIDNTGNANKGVVIFTGA 420
 Db 361 LATINIGAGVATLGGAVIKATTTKLTDDASVLIFTPVVTGAIIDNTGNANKGVVIFTGA 420

Qy 421 STVTDNIGNTAVLAEVSVGAGLLQIQGGVKKANINLTDNASVYFTFGDSTVTSIGGTE 480
 Db 421 STVTDNIGNTAVLAEVSVGAGLLQIQGGVKKANINLTDNASVYFTFGDSTVTSIGGTE 480

Qy 481 LFAVTNIGAGITLRAGGSLAANNIDFGAASNLFEENGAGKNYNLIGTIANGNNATLINA 540
 Db 481 LFAVTNIGAGITLRAGGSLAANNIDFGAASNLFEENGAGKNYNLIGTIANGNNATLINA 540

Qy 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDILDAQAISFKGAASRLFLANVSLQM 600
 Db 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDILDAQAISFKGAASRLFLANVSLQM 600

Qy 601 IELSLKLIIPVLLTVVVS 619
 Db 601 IELSLKLIIPVLLTVVVS 619

RESULT 2
 AAR94014
 ID AAR94014 standard; Protein; 1026 AA.
 XX
 AC AAR94014;
 XX
 DT 21-MAY-1996 (first entry)
 XX
 DE Caulobacter S-layer protein.
 XX
 KW S-layer; rsaA gene; surface protein; biofilm; vaccine.
 XX

OS Caulobacter crescentus strain CB15.
 XX
 FH Key Location/Qualifiers
 FT 860..905
 ET /note= "glycine-aspartic acid repeat region"
 XX
 PN US5500353-A.
 XX
 PD 19-MAR-1996.
 XX
 PF 09-JUN-1992; 92US-0895367.
 XX
 PR 09-FEB-1994; 94US-0194290.
 PR 09-JUN-1992; 92US-0895367.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 XX
 PI Bingle WH, Smit J;
 DR WPI; 1996-171046/17.
 DR N-PSDB; AAT17717.
 XX
 PT Expression and presentation of polypeptide heterologous to
 PT Caulobacter S-layer protein to bacterium's environment - by cloning
 PT in-frame into Caulobacter rsaA gene, and expressing as fusion prod.
 PT with S-layer protein
 XX
 PS Example 2; Fig 6a-c; 22pp; English.
 XX
 CC The Caulobacter crescentus CB15 paracrystalline S-layer protein
 CC (AAR17717) is encoded by the rsaA gene (AAR17717). It can be used as
 CC a fusion partner with polypeptides of interest, allowing
 CC presentation of the polypeptide on the surface of the Caulobacter
 CC cells. The bacterium is cultured as a biofilm in a bioreactor
 CC or may be used to present an antigenic epitope (see AAR94016 and
 CC AAR94018) to the environment e.g. for use as a vaccine.
 XX
 SQ Sequence 1026 AA;

Query Match 11.0%; Score 333.5; DB 17; Length 1026;
 Best Local Similarity 24.7%; Pred. No. 2.9e-13;
 Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

Qy 16 GLKTALETTSTAAILMTGSGVL-----GAARTVTAD-GAELAAGT-----NIGP 58
 Db 315 GIETMNVTSAAITLNTSSGVTGLTALNTNTSGAAQVTAGAGONLTATTAOAAANNVAV 374
 Qy 59 GAGAFV-----AGSTLOYTGFTVTDADVSRALDNNFAAGLFSVTGDISL----- 105
 Db 375 DGGANVTAVSTGVTSGTFT--VGANSAASGTVSVSVANSSTTTTGATVATGTAATVTAQT 432
 Qy 106 -GSYVD-----TGGANKLAVNIDDLGLTLTLTGTGTAAYGANPALLFOGGQAAANN 154
 Db 433 AGNAVNTLTQADVTVTGNSTTAVTVQ--TAAATAGATVAGVNGAVTITDSAAASAT 490
 Qy 155 TYTALGNITLGGANAGLTIASDPDVLGPITLAG-----NIDG 191
 Db 491 TAGKIATVTLGSFGAA-TI--DSSALTTVNLSGTGSLGIGRGALTATPANTLTLLNVNG 547
 Qy 192 ---GGIITDNTDAA-----INGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNA 240
 Db 548 LTTTGALTDEAAADGGFTTINAGSTIASLVAADATTNLNISGDARVTTTSHATAA 607
 Qy 241 APVLTITNA-----NAVLTGAVDNT----- 260
 Db 608 LTGITVNSVGATLGAELATGLVFTGAGRDSILLGATTKAIVNGAGDDTVTVSSATLGA 667
 Qy 261 ----TGSD--DVGVNLNGALSQVTGNIGNTSLATISVGAGTATLG-----G 302
 Db 668 GGSVNGGSDGTDVLVANVNGSSFSADPAFGGFETLRV---AGAAQGSNHNANGFTALQLG 723
 Qy 303 AVIKATTTKLTNA-----SVLTITNAVLTGAVDNTTGGDVGNNVNLGGALSQVTGNIGN 357
 Db 303 AVIKATTTKLTNA-----SVLTITNAVLTGAVDNTTGGDVGNNVNLGGALSQVTGNIGN 357

Db 724 ATAGATT--FTNVAVNGLTVLAAPTGTTTVTILANATGSDVFNLTSSSAAALACTVA- 780
 QY 358 TNSLATINIGA---GVATLDGAVIKATTTKLTDDASVLFTNPVVTGCAIDNTGNANKG 413
 Db 781 LAGVETVNIATDNTNTAHVDLTTLQATS- 820
 QY 414 VVIFTGASTVTDNIGNTAVLAESVAGLLQIQGGVVKANALNTDNASVVTFTGDSVTI 473
 Db 821 LNL-----INTGNTA-----VTSFASAVTGTGSVAVFVSANTIV 855
 QY 474 GSIGGTELEPATVNIAGITLRAGGSLAANNIDFGA-ASNLEFNGPAGKYNLIGTIANGN 532
 Db 856 GEV-----VTIRGGAGADSLTGSATANDTIIGGAGADTLVVTGTT-----DFTTGT 902
 QY 533 NATL-NINAGT-----VIANDVSGITVAQINQNNKIFVINAKNADVDILDAQAISFKA 587
 Db 903 GADIFDINAIGTSTAFVTITDAVG-----DKLDLVGISTNGA 940
 QY 588 -ASRLFLANVSL 598
 Db 941 IADGAFGAATL 952

RESULT 3

AAW37490
 ID AAW37490 standard; Protein; 1026 AA.

AC AAW37490;

XX 20-APR-1998 (first entry)

XX Caulobacter crescentus S-layer rsaa protein.

XX S-layer; rsaa gene; Caulobacter; vaccine; antigenic; ligand; enzyme;
 KW metallochionein; heavy metal; water; sewage; xylanase; cellulase;
 KW wood pulping.

XX Caulobacter crescentus.

XX WO9734000-A1.

PD 18-SEP-1997.

XX 10-MAR-1997; 97WO-CA00167.

XX 12-MAR-1996; 96US-0614377.

XX (UYBR-) UNIV BRITISH COLUMBIA.

XX Bingle WH, Nommellini JF, Smit J;

XX WPI: 1997-470880/43.

DR N-PSDB; AAY01866.

XX New DNA containing sequence for C-terminal region of Caulobacter
 PT S-layer protein - expressed as fusion proteins containing antigenic
 PT peptides in Caulobacter, useful as live vaccines

PS Example 2; Fig 6; 58pp; English.

XX The present sequence represents a Caulobacter S-layer protein used in an
 CC example of the present invention. A new DNA construct has been developed
 CC which contains at least one restriction site for insertion of DNA
 CC upstream of DNA encoding a C-terminal region of at least the last 82
 CC amino acids (aa) of Caulobacter S-layer protein. Caulobacter containing
 CC the DNA constructs above additionally containing a sequence encoding a
 CC heterologous polypeptide, are particularly useful in live vaccines
 CC (where the heterologous polypeptide is an antigen). They can also be
 CC used for production of e.g. ligands, enzymes or other proteins, e.g.
 CC metallochioneins to remove heavy metals from water or sewage, or
 CC xylanase or cellulase for use in wood-pulping. All known Caulobacter
 CC strains are harmless, and stable in outdoor environments, including
 CC water (so suitable for vaccinating fish) or soil. They are well suited

CC for growing in biofilm reactors and produce S-layer proteins, which is
 CC an ideal system for presentation of antigens, at high level.
 XX Sequence 1026 AA;
 SQ Query Match 11.0%; Score 331; DB 18; Length 1026;
 Best Local Similarity 27.3%; Pred. No. 4.2e-13;
 Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;
 QY 23 TTSTRAIMLTGSGVGAART-----VTADG-----AELAGATNIGPGA 60
 Db 413 TTTTGAIAVGTGTAHTVAOTAGNAVNTLTQADVTVTGNSSTTAVTQTAAAT-----A 467
 QY 61 GAFVAGSTLOYTCAFTVTDADVS-----VRALDLNNEAAGL-----FSVTG-- 101
 Db 468 GATVAG---RVNGAVTITDASAASATTAGKIATVTLTGSGAATIDSSALTNNLSGTGS 524
 QY 102 -DISLGSVVDTTGGANKLAVNIDDLTLTGTGTAYGANPALLFOGGOAANNNTYALG 160
 Db 525 LGIGRGALTATPTANTLTNLVNG---LTTTGALT-----DSEAAADDGFT-- 566
 QY 161 NITLGGANAGLTIAS-----DPDVLGPITLAGNIDGG--IITDNTDAAING-TIGTNP 213
 Db 567 TINIAGSTASTIASLVAADATTL-----NISGDARVTITSTAAALTGITVNSVGA 619
 QY 214 ---AQISIGASTLSLGA---VIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264
 Db 620 TLGAELATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTIVTSSAIL-GAGGSVNGDG 676
 QY 265 -DVGVNLNGALSQVGTGNIGTNSLATISVGAGATATLG-----GAVIKATTTK 311
 Db 677 TDVLVANVNGSFSADPAFGGFETLRV---AGAAQGGSHNANGFTALQLGATAGATT-- 730
 QY 312 LNNAA-----SVLTTLNAVLTGAVDNTTGGDNNVGVNLSGALSQVGTGNIGTNSLATINI 366
 Db 731 FTNVAVNGLTVLAAPTGTTTVTILANATGSDVFNLTSSSAAALACTVA-LAGVETVNI 789
 QY 367 GA---GVATLDGAVIKATTTKLTDDASVLFTNPVVTGTAIDNTGNANKGVVIFTGAST 422
 Db 790 AATDNTTAHVDLTTLQATS- 823
 QY 423 VTDNIGNTAVLAESVSVGAGLLQIQGGVVKANALNTDNASVVTFTGDSVTGSGTSELF 482
 Db 824 --TNTGNTA-----VTSFASAVTGTAPAVTFVSANTVGEV----- 858
 QY 483 ATVNIGAGITLRAGGSLAANNIDFGA-ASNLEFNGPAGKYNLIGTIANGNATL-NINA 540
 Db 859 VTIRGGAGADSLTGSATANDTIIGGAGADTLVVTGTT-----DFTTGTGADIFDINA 911
 QY 541 AGT---VIANDVSGITVAQINQNNKIFVINAKNADVDILDAQAISFKA-ASRLFLAN 595
 Db 912 IGSTFAVVTITDAVG-----DKLDLVGISTNGAIDAGAFGA 949
 QY 596 VSL 598
 Db 950 VTL 952
 RESULT 4
 ID AAY44757
 ID AAY44757 standard; Protein; 1026 AA.
 XX AAY44757;
 XX 04-MAY-2000 (first entry)
 DT Caulobacter crescentus surface layer protein.
 DE Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
 KW recombinant fusion protein cleavage; enzyme; protein polymer;
 KW antibacterial enzyme; foodstuff.
 XX Caulobacter crescentus.
 OS


```

Db 990 GSIGGNAAV--TLNLTGALTYKSSINANS---GTLVINAKDAELNGEASGHNTTV 1044
QY 526 GTIANGNATLNINAAAGTVIANDVSTGTVAQINIQNNKIFIVINAKNADVDILDAQAISPK 585
Db 1045 -----NAT-NANGSGVIAT-----TSSRVNITGD-LITINGLN-----IISK 1080
QY 586 GAASRLFLANVSLQMIELSLLKIYVLLTV 616
Db 1081 NGINTVLLKGVKID-----VKYIQGVASV 1105

```

RESULT 8

```

AAB01830
ID AAB01830 standard; Protein; 1222 AA.
AC AAB01830;
XX
DT 11-SEP-2000 (first entry)
DE H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
OS Haemophilus influenzae strain K1.
XX
FH Key Location/Qualifiers
FT Misc-difference 307
FT /note= "Encoded by GG"
FT
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
DR WPI: 2000-303789/26.
DR N-PSDB; AAA52180.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8: Fig 20A-R; 307pp; English.
XX

```

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the γ promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae strains J9YC, K1, K21, LCDC22, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or

```

CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1222 AA;

```

```

Query Match 10.5%; Score 316.5; DB 21; Length 1222;
Best Local Similarity 23.8%; Pred. No. 4.4e-12;
Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

QY 42 TVTADGAELAAAGTNIIGPGAGAF-VAGSTLQYTGAFVTADVSVRALDLNN-FAAGLFSV 99
Db 536 TNGNDNTEIQIGNISQKEGNTLTISSDKVNITERIFI-KAGVNGDSDSNEATSANLTIK 594
QY 100 TGDISLGSVVDVGGANKLAVNIDDLGLTLTGTGTAAYCANPALLFQGGQAANNVYTTAL 159
Db 595 TKELKLTNDLNTSGFNKAEITAKDNSNLT-----GDSNAGNTDAK- 636
QY 160 GNITLGGANAGLTIASDPDVLGPITLAGNIDGGIITDNDAAINGTNTNPAAOISIG 219
Db 637 -KVTFSNVSKISASDHNV-----TLNSKVETSG-DTSDTEGNNNTGLITAKNVTYN 690
QY 220 ASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALSOVT 279
Db 691 NNITSHKVTNITASENVTTKAGTTINATTGSVEVTAKTGDIKGGIESGNVNITASGDT 750
QY 280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTTKLTNAASVLTNAN-----LTGAVD 331
Db 751 LNVSNITGQNTVAAASGAVTTTKGSTINAT-----TGNANITTKTGEINGEYKASGNV 806
QY 332 NTTGGDNVGVNLSGALSOVTGNIGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASV 391
Db 807 ITASGNTLVNSNITGQNTVTANS-----AITTEGSTINAT-----TGDANI 850
QY 392 LIFTNPVVVTGAIDNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVVGAGLLQIQ--- 446
Db 851 -----TTQTGNGINGKVESSGSVTLIATGQTLAVGNISGDTV--TITADKGLTQTSS 902
QY 447 -----GGVVKANALNTDNASVVTFTGD-----STVTSIGG 478
Db 903 KINGTKSVTSSQSGDISGTISGNTVSVSATSLTQTQAGSKIEAKTGEANVTSATGTIGG 962
QY 479 TELFATVNI-----GAGITLRA--GGSIAANNIDFGAASNLEFNGPAG 519
Db 963 TISGNTVNVNTANDNLTIKDGARIKATGAVTLTATGGLTGTETSSDITSSNGQTLTA- 1021
QY 520 KYNVLTIGTIANGNNATLNINAAAGTVIAN---DVSIGTVAQINIQNNKIFIVINAKNADVD- 575
Db 1022 KDSSTAGSI-NAANVLTNTTGLTTLTVAGSKIEAASGTL-----VINAKQAQLDG 1069
QY 576 -----ILDAQAISPKGA--ASRLFLANVSLQMIELSLLKII 609
Db 1070 AASGDHTVVVNATNANGSGSVIATTSRVNITGLITINGLNII 1112

```

RESULT 9

```

AAB01828
ID AAB01828 standard; Protein; 1228 AA.
AC AAB01828;
XX
DT 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain K1 full-length HMW1A protein, SEQ ID NO:34.
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW

```

recombinant production; Escherichia coli; antibacterial; vaccine; human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.

Haemophilus influenzae strain K1.

Key Location/Qualifiers
Misc-difference 313 /note= "Encoded by GG"

W0200020609-A2.
13-APR-2000.
07-OCT-1999; 99WO-CA00938.
07-OCT-1998; 98US-0167568.
08-DEC-1998; 98US-0206942.
(CONN-) CONNAUGHT LAB LTD.
Loosmore SM, Yang Y, Klein MH;
WPI: 2000-303789/26.
N-PSDB; AAA52179.

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -
Claim 12; Fig 20A-R; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwaBC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwiABC and hmwiABC. Each hmwaBC operon comprises hmwa, hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins and the hmwb and hmwc genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwaBC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents an HMWA protein from a non-typeable strain of H. influenzae.

Sequence 1228 AA;
Query Match 10.58; Score 316.5; DB 21; Length 1228;
Best Local Similarity 23.88; Pred. No. 4.4e-12;
Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

42 TVTADGAELAAAGTNGFAGAF-VAGSTLQVTGATFTDADSVRALDLNN-FAAGLFSV 99
542 TNGDNTETIQGGNISKEGNTLTSSDKVNITERITTI-KACVNGDSDSNEATSAITIK 600
100 TGDISLGVSDTGGANKLAVNIDPLTTLTGCTGAAYGANPALLFQGGAAANNNTYAL 159
601 TKELKLTNDLINSFENKAEITAKDNSNLT-----GDNSDAGNTDAK- 642

160 GNITLGCANAGLTATSDPDVLGPTTLAGNIDGGIITDNTDAINGTIGNTPAAQISIG 219
643 -KVTFNVKDSKISASHNV-----TLNSKVETSG-DTDSLEDGNNNTGLTITAKNVTVN 696
220 ASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDVCVNLNGLALSQVT 279
697 NNITSHKVTNITASENVTTRAGTINATTSVETAKTGDIKGGIESNGVNITASGDT 756
280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTTKLTNAASVLTLTNAV-----LTGAVID 331
757 LNVSNITQNVTVAAASGAVTTTKGSTINAT-----TGNANITTKTGEINGEVKSASNVN 812
332 NTTGGDNVGVNLSGALSQVVTGNIGNTNSLATINIGAGVATLDCAVIKATTTKLTDDASV 391
813 ITASGNTLNVSNITQNVTVTANS-----AITTEGSTINAT-----TGDANI 856
392 LIETNPVVVTGAIDNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVSGAGLLQIQ--- 446
857 -----TTQGNINGKVESGSLTATGOTLAVGNISGDTV--TTADKGLKLTQTSS 908
447 -----GGVVKANAILTNDNASVVTFTGD-----STVTGSIIG 478
909 KINGTKSVTTSSQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVTSATGTIGG 968
479 TELFATVNI-----GAGITLRA-GGSLAANNIDFGAASNLFEENGPA 519
969 TISGNTVNVNTANTDNLTKDGARIKATGGAVTLTATGTLTSTSSITSSNGQTTLTA- 1027
520 KYNLIGTIANNNATINAAAGTVIAN--DVSIGTVAQINQNNKIFVINAKNADVD- 575
1028 KDSIAGSI-NAANVTLTGTLTVAGSKIEAASGTL-----VINAKDAQLDG 1075
576 -----ILDAQATISFKGA--ASRLFLANVSLQMIELSLKII 609
1076 AASGDHTVWATNANGSGSVIATTSRRVNTIGDLITINGLNII 1118

RESULT 10
AAB01835
ID AAB01835 standard; Protein; 1095 AA.
XX AAB01835;
XX AC (first entry)
DT 11-SEP-2000 (first entry)
XX Haemophilus influenzae strain LCDC2 mature HMWA protein, SEQ ID NO:45.
DE Haemophilus influenzae strain LCDC2.
KW Mature HMW protein; hmw gene; hmwa1; hmwa2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX Haemophilus influenzae strain LCDC2.
OS W0200020609-A2.
PN 13-APR-2000.
XX 07-OCT-1999; 99WO-CA00938.
XX 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
XX WPI: 2000-303789/26.
DR N-PSDB; AAA52184.
XX Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide

protection against Haemophilus induced diseases in humans -

Claim 8; Fig 22A-P; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwa, hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins and the hmwb and hmwc genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Jovic, K1, K21, LDC22, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMWA protein from a non-typeable strain of H. influenzae.

Sequence 1095 AA;

Query Match 10.38; Score 310; DB 21; Length 1095;

Best Local Similarity 23.18; Pred. No. 1e-11;

Matches 174; Conservative 96; Mismatches 240; Indels 242; Gaps 35;

Qy	18	KTALFTTTAAIMLTGSGV-----LGAARTVTADGAELAAGTNIGPGAGA	62
Dd	350	RTAFIESDITLATNGNISLNOVAGIDNLOKSLVANKNIITEGNTIPLAADKKP----	405
Qy	63	FVAGSTLOYTGAFVTID-ADVSVRALDDLNNRAGLFESVTGDI-SLGVSVDYTGKANGLAVN	120
Dd	406	-----IEIKGNITVEGANVTILRSANYGNDKSAL-SIRGNTVNKGNLTVTGS-----AIN	454
Qy	121	IDDLGLTITLTGTGAAYGANPALLOFQGAAAANNVTYALCNTTLGGAN-----AGLTI	173
Dd	455	IENLIV-----EGSAKFLANPNYSFNVSGLFDNQOKSNI-SIAKGGAEHFKDINNTKSLNI	509
Qy	174	ASDPDWLGPTTLAGNI---DGGGIIIPDNTDAA---INGTI-----GN-----TNPAAQI	216
Dd	510	TTNSDSAYRTIIEBIGNITSNGDLNITDKNNBAEIQIGGNIQKEGNLTISSDKINITNOI	569
Qy	217	SI-----GASTISLGAVIKATTKLT-----T-----	238
Dd	570	TIKKGVNKEDSDSTANNANLTIKTKELOITGDNLINSGFDAEITAKEGADLIIGNSDNN	629
Qy	239	---NAAPV-----LTLITNANAVLTGAVDNITCGDDVGVLNUNGA-----LSQVTGN	281
Dd	630	NNANAKVTNPQVKDSKISADSHNVTNLNSKVETSNNGDAESNNGDTSLTTNAKNITVN	689
Qy	282	IGNTNSLATISVGA--GTATFLGAVIKATTKLTUNAASVLITLNVLITGAVDNTTG----	335
Dd	690	-NNITSHKTWNITAENVITKAGTTINATTGSVE-----VTAKTGDIKKVESTSGSVTL	743
Qy	336	---GDNVGVVNLSCALSQVVTGNIGTNSLATINIGAVATLDGAVIKATTKLTDDASVL	392
Dd	744	TATGEALAVSNI SGNTVTITANKGK----LTTQAGSTVSAING---VTASSQSGDISGT	795
Qy	393	IFTNPVVVTGAI DNTGANKGVVI FTGASTVTDNICNTAVLAESVGVAGLLQIOGGVKA	452
Dd	796	ISGNTKVASI GBLTTKSGSEIKAKTEGANVTSATCTI-----GGTISG	839

	QY	453	NAINLTDNASVVTFTGDSTV-----	472
	Dd	840	NNVNVTAN-----TGDLTVDEADAADKATGGAATLTATSGKLTTKASSITTSANNQNVNLS	893
	QY	473	--TGSIGGTLEFA--TWNIAGGITLRAGSLAANNIDFGA---ASNLFENGPAKGKNYNL	524
	Dd	894	AKDGSIGNINAAENVTLNTTGALTUVKGGSSINANS---GLVLNAKADEUNGESGNHTV	950
	QY	525	IGTIANGNNATLNINAAGTVIANDVSIGTVAQIONQNKFVINAKNADVILDQAIAISF	584
	Dd	951	V-----NAT-NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----IIS	986
	QY	585	KGARSRFLVANSLQMIEUSLLKIIVPLLV	616
	Dd	987	KNGINTVLLGVKID-----VKYIQPGIASV	1012
		RESULT 11		
		AABO1834.	ID AABO1834 standard; Protein; I101 AA.	
	XX	AC	AABO1834;	
	XX	DT	I1-SEP-2000 (first entry)	
	XX	DE	Haemophilus influenzae strain LCDC2 HMWLA protein, SEQ ID NO:43.	
	XX	KW	HMW protein; hmw gene; hmwa1; hmwA2; high molecular weight; non-typeable Haemophilus influenzae; NTHi; non-encapsulated; recombinant production; Escherichia coli; antibacterial; vaccine; human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.	
	OS	HM	Haemophilus influenzae strain LCDC2.	
	PN	WO	2000020609-A2.	
	PD	13-	APR-2000.	
	PF	07-OCT-	1999; 99WO-CA00938.	
	PR	07-OCT-	1998; 98US-0167568.	
	PR	08-DEC-	1998; 98US-0206942.	
	PP	(CONN-) CONNAUGHT LAB LTD.		
	PI	Loomore SM, Yang Y, Klein MH;		
	DR	WIPI; 2000-303789/26.		
	DR	N-PSDB; AAA52183.		
	PT	Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans - Claim 12; Fig 22A-P; 307pp; English. The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw genes, clusters termed hmwlabc and hmwa2abc. Each hmwABC operon comprises hmwa, hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins and the hmwb and hmwc genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAA52175-AA52198) and HMWA proteins (AABO1824-BO1849) from the non-typeable H. influenzae strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW		

168 N ---AGLTI-----ASDPDLGPITLAGNIDGGII-----TDNTDAINGT 206
 169 NIICAGGCTTQICANGCTIKLTSTNNIVYDFLAIATDGTGVVDASSLTNAOTLTINGK 228
 207 IGWNPAAQISIGASTLSLGGAVIKATTKLTNAAPVLTITNANAVLTGAVDNTT-----261
 229 IH-----TIGANKTLGTGTIGSKTSLSGNVAIVLIGNDGAAGVAFANDTTLITRT 280
 262 ---GGDV--GVNLNGLSALSOVTGNIGN-TNSLA-----TISVAGT-----297
 281 TNAAGGKIIFXPVN-NGTLAAGT-NGSATNPLAGINTSGKGVVDVTLNVEGCVNLYA 338
 298 -----ATLGGAVIKATTK-----LTNAASVLTITNNAVLTG-----328
 339 TNITTTDANVGSFVFNASSNIYSGTVGGQGNKFNTVALENGTTVKFLGNATFNGNTTI 398
 329 -----AVDNTGGD--NVGVNLS-----GALSOVT-----352
 399 AANSTLOISGNYTADCVASADGTGVEFVNTGPITVTLNKEAAPVNALKOITVSGPENVV 458
 353 ---GNIGNTNSLAT-----INIGAVATLDG-----AVIKAT-----TTKLTLD 387
 459 INEIGNAGNNGAVTDTIAENSSIGAVVFLPGIPENDAGNTMPLIKSTVGNTAKGPD 518
 388 -----DASVLTNPVVVTGAIDNTGNANKGVVIFTGASTVTDNIGNTAVL 433
 519 VPSVVLGVDSVADGQVIVDQNNIVGIGLSDNGIIVNATTLYAGISTLNN-----571
 434 AEYSVAGLLOIOGGV--VKANAINLTDNASVVTFTGDSVTGTSIGGTLPAT-VNIGAG 490
 572 ---OGVTLSGGVPNTPGTVYGLSTISASKFKGVFTTGTGYNLSNIATNATINDG 625
 491 ITRAGSLAANNIDFGAENLEFNGPAGKNYNLIGTIANGNATNLINAAAGTVIANDVS 550
 626 VTTTGGI-----AGIGIGFDGKI-----TLGSV-NGNG---NVPFADGILSNST 667
 551 I-----GTVAQIQNNKIFVINAKNADVDILDQAISFKGAASRLFL-ANVSLQMI 601
 668 MICTKANNGTVTYL-----GNAFVGNIGSDTPV---ASVRFETGSGAGLOGNIVSQVI 720
 602 E 602
 721 D 721
 RESULT 13
 AAR41731
 ID AAR41731 standard; Protein; 1338 AA.
 XX
 AC AAR41731;
 XX
 DT 26-APR-1994 (first entry)
 XX
 DE High molecular weight protein 3 (HMW3).
 XX
 KW HMW; high molecular weight protein; virus; vaccine; influenza;
 KW epitope; immunity; haemophilus influenzae.
 XX
 OS Haemophilus influenzae.
 XX
 PN W09319090-A.
 XX
 PD 30-SEP-1993.
 XX
 PF 16-MAR-1993; 93WO-US02166.
 XX
 PR 16-MAR-1992; 92GB-0005704.
 XX
 XX (BARE/) BARENKAMP S J.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Barenkamp SJ;
 XX

DR WPI: 1993-320683/40.
 DR N-PSDB; A049510.
 XX High molecular weight surface proteins - of non-typeable
 PT haemophilus which exhibit immunogenic properties
 XX Claim 5; Figure 10; 100pp; English.
 XX The isolation and purification of the high molecular weight protein
 CC enables the identification of the major protective epitopes of the
 CC protein by conventional epitope mapping. These epitopes can then be
 CC synthesised using standard techniques and incorporated into fully
 CC synthetic or recombinant vaccines.
 XX Sequence 1338 AA;
 SQ
 Query Match 9.9%; Score 299.5; DB 14; Length 1338;
 Best Local Similarity 22.8%; Pred. No. 6e-11;
 Matches 166; Conservative 98; Mismatches 247; Indels 217; Gaps 36;
 QY 19 TALFTTSTAAML---TSGVLG-RARTVTADGAELAACT-NIGPCAGA--FVAGSTLQ 70
 Db 410 TTLNVTSGSKFNLSIDSTGSGTSPSIRNAELNGITFNKATFNIAOGSTANFSAKASIMP 469
 QY 71 Y--TGATFTVDADVRA-----LDLNNFAAGL-----FSVTGDISLGSVVDTG 112
 Db 470 FKNANYALFENEDISVGGGVNFKLNASSNIQTGCVIILKSNFNVSGGSTLN--LKA 527
 QY 113 GANKLAVNIDDDGLTLTLTG-----TGT-----AAYGANPALLFQGGQ---AAANNTY 156
 Db 528 GSTETAFSIENDLNLTATGNNITIROVEGDSRVNKGVAAKKNITFKGNNITFGSOKATT 587
 QY 157 TALCNTLTG-GANAGLTIASDPDVLGPITLAGNIDGGIITDNDRAINGTIGN-----209
 Db 588 EIKGNTINKNTNLTARGANFAENKPSLNAGNVINNGNL--TAGSIINAGNLTVSKG 645
 QY 210 -----TNPAQI-----SIGASTLSL--GGAVIKATTKLTNAAPVLTITNAN---AV 252
 Db 646 ANLQAITNYTFNVAGSFDNNGASNIARGAKFK---DINTSSLTNTNSDTYRTI 701
 QY 253 LTCAVDNTTGGDDVGVLN-----LNGALSQVOTNIG-----NTNSLATISVGA---295
 Db 702 IKNISNKG--DLNIIDKSDAEIQIGGNISOKEGNTLISDKVNITNOITIKAGVEGG 759
 QY 296 -----GATLGGAVIKATT-309
 Db 760 RSDSSAEANALTIQIKELKLAGDLNLSGNPKAEITAKNGSDLTIGNASGNNADAKRVTF 819
 QY 310 -----TKLTNAASVLTITNAVLTAVDNTTGGDNVGVNLSGALSQVOTNIGNTNSLATI 364
 Db 820 DKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISA--KQVTN-NVVTSHKTI 876
 QY 365 NIGA---GVATLDGAVIKATTKTL-----TDDASVLTFTNPVVTGAIDN-406
 Db 877 NISAAAGNVTTKEGTTINATTSVEVTAQNGTIKGNITSONVTATENLVTTENAVINA 936
 QY 407 -----TGNANKGVVIFTGASTVTDNIGNTAVLAE-----VSVGAG-LLOIQGG 438
 Db 937 TSGTVNISTKTGDIKGIESTSCNVNITAS-GNTLKVSNITGQDVTVTADAGALTITAGS 995
 QY 449 VKKANAINLTDNASVVFTGD---STVTGSIIGGTETFPATNIGAGITLRAGGSAAANNIDF 506
 Db 996 TISAT-----TGNANITTKTGDKINGKVESSSGVTLVAT-----GATLAVGNI--1038
 QY 507 GAASNLEFNGPAGKNYNLIGTIANGNATNLINAAAG-----TVIANDVSIGT 553
 Db 1039 -SCNTVTIITADSGKLTSTVGSTINGNTSVTSSQSGDIEGTISGNTVNTVNTASTGDLTIGN 1097
 QY 554 VAQINIQN 561
 Db 1098 SAKVEAKN 1105

RESULT 14

AAW30291 standard; Protein: 1598 AA.

AAW30291;

14-APR-1998 (first entry)

Non-typeable Haemophilus high mol.wt. surface protein HMW3.

Non-typeable Haemophilus; high molecular weight surface protein; HMW3; immunogen; vaccine; otitis media.

Haemophilus influenzae strain 5.

Key Location/Qualifiers

Misc-difference 113 /note= "encoded by GTG"

Misc-difference 864 /note= "encoded by TGT"

W09736914-A1.

09-OCT-1997.

01-APR-1997; 97WO-US04707.

01-APR-1996; 96US-0617697.

(BARE/) BARENKAMP S J.

Barenkamp SJ;

WPI: 1997-503038/46.

N-PSDB: AAT90992.

High molecular weight proteins of non-typeable Haemophilus influenzae - useful for vaccine production

Claim 1; Page 93-97; 183pp; English.

This protein comprises the high molecular weight surface protein HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that has the immunological ability to protect against disease caused by a non-typeable Haemophilus strain and is characterised by at least one surface-exposed B-cell epitope that is recognised by monoclonal antibody AD6. The HMW3 amino acid sequence was deduced from an isolated hmw3 gene (see AAT90992). HMW1 (see AAW30293), HMW2 (see AAW30294) and HMW4 (see AAW30292) have also been identified. A conjugate comprising HMW3 linked to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150 amino acids corresponding to at least protective epitope of HMW3 are also claimed. HMW proteins, conjugates and peptides can be used in vaccines, as immunogens for preparation of antibodies and as antigens for detection of these antibodies.

Sequence 1598 AA;

Query Match 9.9%; Score 299.5; DB 18; Length 1598;

Best Local Similarity 22.8%; Pred. No. 7.4e-11;

Matches 166; Conservative 98; Mismatches 247; Indels 217; Gaps 36;

19 TALFTTSTAAML-----TGSGVLG-AARTVADGAELAAAGT-NIGPGAGA--FVAGSTPLQ 70

670 TLANVTSKFNLSIDSTGSGTSPSIRNAELGTFNKATFNAGSTANFSIKASIMP 729

71 Y--TCAFTVTADYSVRA-----LDLNNFAAGL-----FSVTGDISLGSVDYTG 112

730 FKSNNYALFNEDISVGGGSLNFKLNASSNIOTPGVIKSNQFNVSQGGSTLN--LKAE 787

113 GANKLAVNIDGLTITLTG-----TGT-----AAYGANPALIFOGGQ---AAANNY 156

788 GSTETAFSIENDLNLNATGGNITITROVEGTSRNVKNVGAACKNITFGGNTFGSQKATT 847

QY 157 TALGNITLG-GANAGLTIASDPDVLGPITLAGNIDGGIITDNTDAAINCTIGN----- 209
 DB 848 EIKGNVTINKNTNATLRGANFAENKSPNLTAGNVINNGNLT--TAGSTINIAGNLTVSKG 905
 QY 210 -----TNPAAQI-----SIGASTLSL--GGAVIKATTTKLTNAAPVLTLTNAN---AV 252
 DB 906 ANLOAITNTYTFNVAGSFDNNGASNISIRGCAKPK----DINNTSSLNITNSDTTYRTI 961
 QY 253 LTGAVDNTTGGDDVGVLN-----LNGALSOVVTGNIG-----NTNLSATISVGA--- 295
 DB 962 IKNISNKSQ--DLNIIIDKKSDBAIEIQGNISQKGNLTISSDKVNTNQTITKAGVEGG 1019
 QY 296 -----GATLGGAVIKATT-- 309
 DB 1020 RSDSSEAEANLTIQTKELAGDLNISGFNKAIEITAKNGSDLTIGNAGSGNADAKKVT 1079
 QY 310 -----TKLTNAASVLTLTNAVLTCGAVDNTTGGDNGVNVNLSGALSQVVTGNIGNTNSLATI 364
 DB 1080 DKVRDSKISTDGHNVTLNSEVKTNGSSNAGNDNSTGLITISA--KDVTVN--NNTVSHKTI 1136
 QY 365 NIGA---GVATLDGAVIKATTKL-----TDDASVLIFTNPVVVTGAIDN- 406
 DB 1137 NISAAAGNVITKEGTTNATGTSVEVTAQNGTIKGNITSQNVTVTATENLVTENAVINA 1196
 QY 407 -----TGNANKGVVIFTGASTVTDNIGNITAVLAE-----VSVGAG--LLQIQGG 448
 DB 1197 TSGTVNISTKTGDIKGGIESTSGNVNITAS-GNTLKVSNTIGQDVTVTADAGALTTTAGS 1255
 QY 449 VVKANAINLTDNASVVTFTGD--STVTSIGGTGTELFATVNVIGAGITLRAGSGLAANNIDF 506
 DB 1256 TISAT-----TGNANITTKGDINKGVSSGSVTLVAT-----GATLAVGNI-- 1298
 QY 507 GAASNLEFNGPAGKNYNLIGTFIANGNATLINAAG-----TVIANDVSIGT 553
 DB 1299 -SGVTITADSGKLTSTVGSTINGTNSVTSSOGDIEGTISGNTVNVITASTGDLTIGN 1357
 QY 554 VAQINION 561
 DB 1358 SAKVEAKN 1365

RESULT 15

ABB52855

ID ABB52855 standard; Protein: 2732 AA.

XX AC ABB52855;

XX DT 11-FEB-2002 (first entry)

XX DE Escherichia coli polypeptide SEQ ID NO 1119.

XX KW Escherichia coli; B2/B+A-; antiInflammatory; antibacterial;

XX KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;

XX KW systemic infection; non-diarrhoeal infection; septicaemia;

XX KW pyelonephritis; antibiotic resistance.

XX OS Escherichia coli.

XX PN WO200166572-A2.

XX PD 13-SEP-2001.

XX PF 12-MAR-2001; 2001WO-EP03445.

XX PR 10-MAR-2000; 2000FR-0003145.

XX PR 02-FEB-2001; 2001FR-0001449.

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI: 2001-550253/61.

Job time : 53 secs

xx A library of DNA fragments of Escherichia coli strains for the
PT phylogenetic determination of a given strain comprises polynucleotides of
XX nature B2/D+ A - -

PS Example 6; Fig 6; 646pp; English.

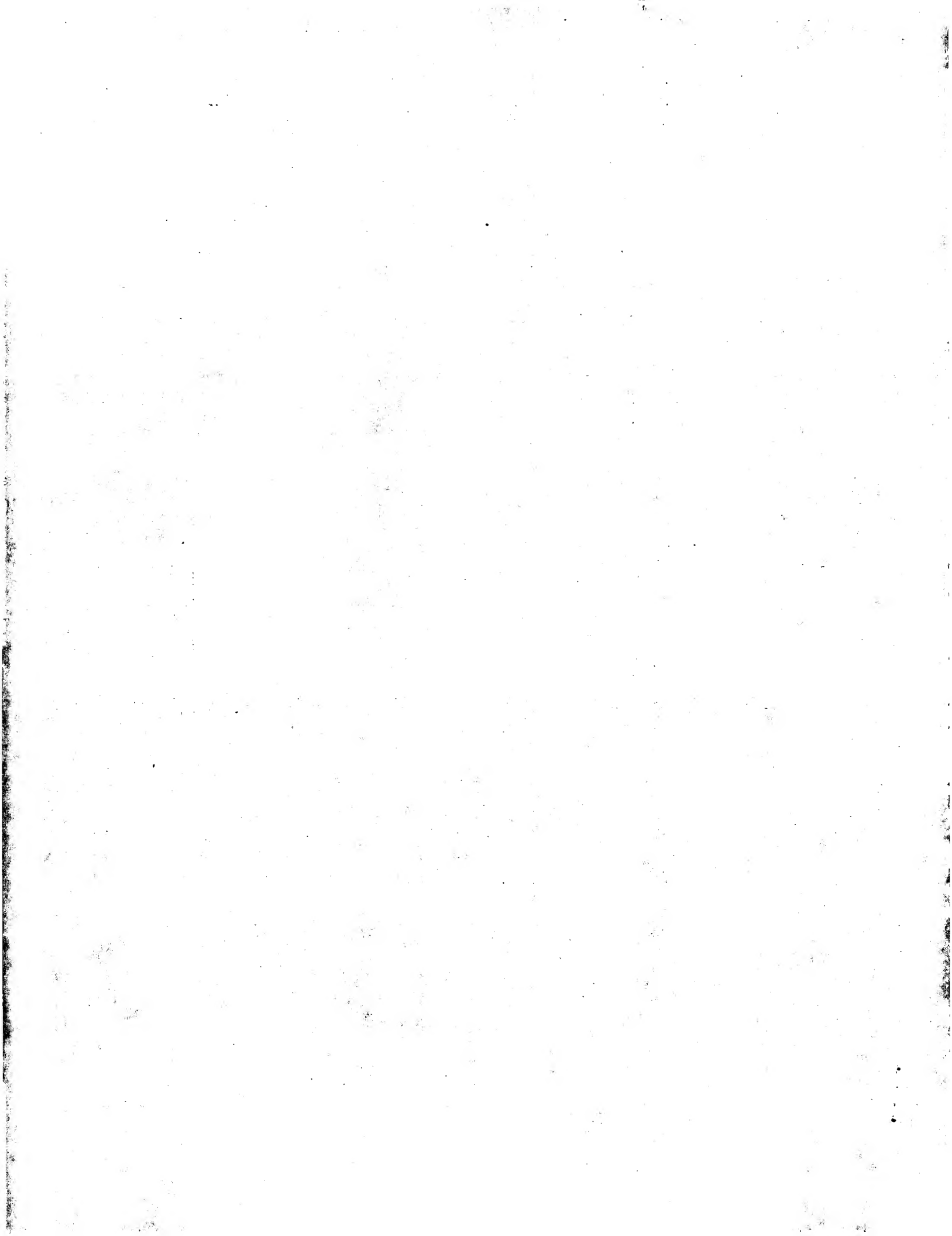
xx The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+ A - - . The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.

XX Sequence 2732 AA;

Query Match 9.9%; Score 299.5; DB 22; Length 2732;
Best Local Similarity 25.6%; Pred. No. 1.4e-10;
Matches 165; Conservative 78; Mismatches 248; Indels 153; Gaps 29;

QY	24	TSTAAMLTGSGVLGAARTVADGAELAGT-----NIGPGAGAFVA-----G 66
Db	791	TSTGIMSGDITLKAA--TTDNRGSLLSAGTSLDGNLSLDNRGTGQGNHVTIRQNSVTNS 849
QY	67	STLOXTGAFVTVDADSVVRALDNNFAGLFSVTGDISL--GSVVDVTGG----- 113
Db	850	GULTGTAALTAAARWASPOPALMNN--GGSLLTSGDLTITAGSITSSGHQWQKRVLTAD 907
QY	114	--ANKLAVNIDGLTITGTCTAAYGANPALLFOGQQAANNYYTALGNITLGGANAGL 171
Db	908	SLANGAIQAADSLTARLTG-----ELVSTAGSKVTSNGEMALSALNL--SNSGQ 955
QY	172	TIASPPDV--LGPITLAGNIDGGIITDNTDAINGTIGNTPAAQISIGASTL----- 223
Db	956	WIAKNLTAKNSLTAGDITG----VDLTIVNQTLNQNQKLLSAGVTLKADSVTN 1011
QY	224	--SLGAVIKATTKLTNA-----APVLTITNANAVLTGAVDNTGGDDVGVNLNGALSO 277
Db	1012	DGQLQGNVTTITAGQLTNGHILQGETLTLT-----ASGVNRRSG----GVLMSRNALNV 1062
QY	278	VTGNIGNTNSLATISVGACTATLGGAVIKATTKLTNAASVLTLTNAVLTGAVDNTGGD 337
Db	1063	STATLSNOSTIOG-----GGVSLNATDRLONDGKILSGSNLTILTAQVLANTGS- 1111
QY	338	NVGVNLSGALSQVTGNIGNTNSLATINIGAV-----ATLDGAVIKATTKLTDDA 389
Db	1112	--GLVQARTLLLDVVYTVNGRVLAT---GSDVKGTTLNNTGTLOGATL---VNYHTFSS 1163
QY	390	SVLIPTNPVV-----TGADNTGN-----ANKGVVIFTGASTV-----TDNI 427
Db	1164	GTLLGTSGLVKRGSSLLQNGTGRLYSAGNLLDADQFSQGVQVATGDVTLKIALALNH 1223
QY	428	GNTAVLAESVGCAGLLQGGVVKANAINLTDNASVTTTGDSTVTGSIIGTTELFATVNI 487
Db	1224	GTLAAGKTLSTVSNATITNGVYMQGDAMVL---GAGEAFTNNGLTAGK--GNSVFSAGR- 1277
QY	488	GAGITLRAGGSL--AANNIDFGAASLEFNGPAGKNYLNIGTIANGNNATLNINAAGTVIA 546
Db	1278	---LFLNAPGSLQGGDVLSNRSRDTISG-----FTGTAGSLTMNVAGTLLN 1322
QY	547	NDVSICTVAQINIQNN--KIFVINAKNADVDIILDAQAISFKGAAS 589
Db	1323	S-----ALIVAGNNKLFTDRLHNQHGDIILAGNSLWVQKDA 1359

Search completed: April 14, 2003, 16:08:54



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:08:04 ; Search time 19 seconds
(without alignments)
1991.748 Million cell updates/sec

Title: US-09-800-065-2
Perfect score: 3022
Sequence: 1 MANISLKFQKAIQKGLKTA.....MIELSLKIIYVLLTVVVS 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 segs, 61136040 residues
Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3022	100.0	619	10	US-09-800-065-2
2	331	11.0	1026	10	US-09-379-931-7
3	303.5	10.0	1600	9	US-10-092-880-10
4	299.5	9.9	1599	9	US-10-092-880-9
5	292.5	9.7	3127	9	US-10-184-644-83
6	286	9.5	1477	9	US-10-092-880-4
7	276	9.1	871	10	US-09-886-468-21
8	271.5	9.1	1536	9	US-10-092-880-2
9	271.5	9.0	1819	9	US-10-184-644-39
10	258	8.5	2340	9	US-10-184-644-379
11	256.5	8.5	943	9	US-09-996-634-131
12	256.5	8.5	943	9	US-09-997-181-131
13	256.5	8.5	943	9	US-09-997-182-131
14	254	8.4	1487	9	US-10-184-644-5
15	253.5	8.4	2397	9	US-10-184-644-29
16	251	8.3	2374	9	US-10-184-644-521
17	249	8.2	2039	9	US-10-192-584-7
18	247.5	8.2	1090	9	US-09-759-130B-134
19	247	8.2	1076	9	US-10-028-072-219

20	247	8.2	1076	9	US-10-121-049-219	Sequence 219, App
21	247	8.2	1076	9	US-10-123-904-219	Sequence 219, App
22	247	8.2	1076	9	US-10-140-470-219	Sequence 219, App
23	247	8.2	1076	9	US-10-175-746-219	Sequence 219, App
24	247	8.2	1076	9	US-10-176-918-219	Sequence 219, App
25	247	8.2	1076	9	US-10-176-921-219	Sequence 219, App
26	247	8.2	1076	9	US-10-137-865-219	Sequence 219, App
27	247	8.2	1076	9	US-10-140-474-219	Sequence 219, App
28	247	8.2	1076	9	US-10-142-431-219	Sequence 219, App
29	247	8.2	1076	9	US-10-143-114-219	Sequence 219, App
30	247	8.2	1076	9	US-10-140-002-219	Sequence 219, App
31	247	8.2	1076	9	US-10-142-419-219	Sequence 219, App
32	247	8.2	1076	9	US-10-123-262-219	Sequence 219, App
33	247	8.2	1076	9	US-10-142-423-219	Sequence 219, App
34	247	8.2	1076	9	US-10-121-050-219	Sequence 219, App
35	247	8.2	1076	9	US-10-141-755-219	Sequence 219, App
36	247	8.2	1076	9	US-10-143-032-219	Sequence 219, App
37	247	8.2	1076	9	US-10-184-644-503	Sequence 503, App
38	246.5	8.2	3150	9	US-10-184-644-81	Sequence 81, App
39	246	8.1	2010	9	US-10-184-644-357	Sequence 357, App
40	244	8.1	1533	9	US-10-184-644-181	Sequence 181, App
41	244	8.1	1786	9	US-10-184-644-481	Sequence 481, App
42	242.5	8.0	1594	9	US-10-184-644-183	Sequence 183, App
43	242.5	8.0	2294	9	US-10-184-644-283	Sequence 283, App
44	239.5	7.9	1723	10	US-09-841-132-394	Sequence 394, App
45	238.5	7.9	1723	10	US-09-841-132-395	Sequence 395, App

ALIGNMENTS

RESULT 1
US-09-800-065-2
; Sequence 2, Application US/09800065
; Patent No. US20020094552A1
; GENERAL INFORMATION:
; APPLICANT: Bouvier, Donald H
; APPLICANT: Crocquet-Valdes, Patricia
; APPLICANT: Stenos, John
; APPLICANT: Walker, David H
; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein
; FILE REFERENCE: 026.00121
; CURRENT APPLICATION NUMBER: US/09/800,065
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: US 60/187,323
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Rickettsia felis
.....US-09-800-065-2

Query Match	100.0%	Score 3022;	DB 10;	Length 619;
Best Local Similarity	100.0%;	Pred. No. 1.9e-185;		
Matches 619;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MANISLKFQKAIQKGLKLTFTTSTAAIMLTGSGVLGAARTVADGAELAGTNGPCA	60	
Db	1	MANISLKFQKAIQKGLKLTFTTSTAAIMLTGSGVLGAARTVADGAELAGTNGPCA	60	
QY	61	GAFVAGSTLQYTGAFVTVDADSVSRALDINNFAAGLFSVTGDISLGSVVDVTGGANKLAVN	120	
Db	61	GAFVAGSTLQYTGAFVTVDADSVSRALDINNFAAGLFSVTGDISLGSVVDVTGGANKLAVN	120	
QY	121	IDDLGLTLTGTCTAAYGANPALLFOGGQAAANNNTYALGNITLGGANAGLTIASDPDVL	180	
Db	121	IDDLGLTLTGTCTAAYGANPALLFOGGQAAANNNTYALGNITLGGANAGLTIASDPDVL	180	
QY	181	GPITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKATTTLTNA	240	
Db	181	GPITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKATTTLTNA	240	

ap 1/25

QY 241 APVLTNANAVLTGAVDNTTGGDDVGVNLNGALSGQVTGNTGNTNSLATISVGAGTATL 300
DB 241 APVLTNANAVLTGAVDNTTGGDDVGVNLNGALSGQVTGNTGNTNSLATISVGAGTATL 300
QY 301 GGAVIKATTTKLTNAASVLTLTNAVLTGAVDNTTGGDNGVNVNLSGALSGQVTGNTGNTNS 360
DB 301 GGAVIKATTTKLTNAASVLTLTNAVLTGAVDNTTGGDNGVNVNLSGALSGQVTGNTGNTNS 360
QY 361 LATINIGAGVATLDGAVIKATTTKLTDDASVLIFPNVVTGAIDNTGNANKGVVIFTGA 420
DB 361 LATINIGAGVATLDGAVIKATTTKLTDDASVLIFPNVVTGAIDNTGNANKGVVIFTGA 420
QY 421 SVTDTNIGNTAVLAESVVGAGLLQIOGGVVKANALNTDNASVVTFTGDSVTGSGIGTE 480
DB 421 SVTDTNIGNTAVLAESVVGAGLLQIOGGVVKANALNTDNASVVTFTGDSVTGSGIGTE 480
QY 481 LFATVNICAGITLRAGGSLAANNIDFGAASNLEFNGPAGKNYNLGTIANGNNATLINA 540
DB 481 LFATVNICAGITLRAGGSLAANNIDFGAASNLEFNGPAGKNYNLGTIANGNNATLINA 540
QY 541 ACTVIANDVSGTGTVAQINQNNKIFVINAKNADVDILDQAQISFKGAASRLFLANVSLQM 600
DB 541 ACTVIANDVSGTGTVAQINQNNKIFVINAKNADVDILDQAQISFKGAASRLFLANVSLQM 600
QY 601 IELSLKIIYPVLLTVVVS 619
DB 601 IELSLKIIYPVLLTVVVS 619

RESULT 2

US-09-379-931-7

; Sequence 7, Application US/09379931

; Patent No. US20020009792A1

; GENERAL INFORMATION:

; APPLICANT: Smit, John

; APPLICANT: Bingle, Wade H.

; APPLICANT: No. US20020009792A1ellini, John F.

; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOBACTER

; CURRENT APPLICATION NUMBER: US/09/379,931

; CURRENT FILING DATE: 1999-08-24

; PRIOR APPLICATION NUMBER: US 08/614,377

; PRIOR FILING DATE: 1996-03-12

; PRIOR APPLICATION NUMBER: US 08/194,290

; PRIOR FILING DATE: 1994-02-09

; PRIOR APPLICATION NUMBER: US 07/895,367

; PRIOR FILING DATE: 1992-06-09

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 1026

; TYPE: PRT

; ORGANISM: Caulobacter crescentus

US-09-379-931-7

Query Match 11.0%; Score 331; DB 10; Length 1026;
Best Local Similarity 27.3%; Pred. No. 1.3e-13;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

QY 23 TTSTAATMLTGSVGLGAART-----VTADG-----AELAAAGTGNIGPGA 60
DB 413 TTTTGAATVGGTAVTVAQTAGNAVNTLTQADVTGTGNSSTTAVTQTAAAT-----A 467
QY 61 GAFVAGSTLQYTGAFVTDADVS-----VRALDNLNFAAGL-----FSVGTG-- 101
DB 468 GATVAG---RVNGAVTITDAAAASATTAGKATVATLGSFGAATIDSSALATVNLSGTGS 524
QY 102 -DISLGSVDVTGGANKLAVNIDGLTTLTGTGTAAVGANPALLFOGQAAANNVTYALG 160
DB 525 LGIGRGALTATPANTLTNLVNG---LTTTGAIT-----DSEAAADGGFT--- 566
QY 161 NITLGGANAGLTIAS-----DPDVLGPITLAGNIDGGG---IITDNTDAAING-TIGTNP 213

DB 567 TINIAGSTASSTIASLVAADATTL-----NISGARVTTTSHATAALGTITVNSVGA 619
QY 214 ---AQISIGASTLSLGGG-----VIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264
DB 620 TLGAEATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTIVTSSATL-GAGGSYNGDGG 676
QY 265 -DVGVLNNGALSGQVTGNTGNTNSLATISVGAGTATLG-----GAVIKATTTK 311
DB 677 TDVLVANVNGSFSADPAFGGFEILRV---AGAAQOSHANGFTALQLGATAGATT-- 730
QY 312 LTNA-----SVLTLTNAVLTGAVDNTTGGDNGVNVNLSGALSGQVTGNTGNTNSLATINI 366
DB 731 FTNVAVNVGLTVLAAPTCTTTTTLANATGTDVFNLTLSAALAAAGTVA-LAGVETVNI 789
QY 367 GA---GVATLDGAVIKATTTKLTDDASVLIFPNVVTGAIDNTGNANKGVVIFTGAST 422
DB 790 AATDTNTTAHVDTTLQATSAK-----SIVVTG-----NAGLNL----- 823
QY 423 VTDNIGNTAVLAESVVGAGLLQIOGGVVKANALNTDNASVVTFTGDSVTGSGIGTELF 482
DB 824 -TNTGNTA-----VTSFASAVTGTAPAVTFVSANTTVGEV----- 858
QY 483 ATVNIGAGITLRAGSLLAANNIDFGA-ASNLEFNGPAGKNYNLGTIANGNNATL-NTNA 540
DB 859 VTIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DTFTGTTGADIFDINA 911
QY 541 AGT-----VIANDVSGTGTVAQINQNNKIFVINAKNADVDILDQAQISFKGA-ASRLFLAN 595
DB 912 IGTSTAFVTITDAVG-----DKLDLVGISTNGAIDAGAFGA 949
QY 596 VSL 598
DB 950 VTL 952

RESULT 3

US-10-092-880-10

; Sequence 10, Application US/10092880

; Patent No. US20020164354A1

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J.

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE

; TITLE OF INVENTION: HAEMOPHILUS

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/10/092,880

; CURRENT FILING DATE: 2002-03-08

; PRIOR APPLICATION NUMBER: 09/155,614

; PRIOR FILING DATE: 1998-09-30

; PRIOR APPLICATION NUMBER: 08/617,697

; PRIOR FILING DATE: 1996-04-01

; PRIOR APPLICATION NUMBER: PCT/US97/04707

; PRIOR FILING DATE: 1997-04-01

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 10

; LENGTH: 1600

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-10-092-880-10

Query Match 10.0%; Score 303.5; DB 9; Length 1600;
Best Local Similarity 24.0%; Pred. No. 1.3e-11;
Matches 163; Conservative 89; Mismatches 249; Indels 177; Gaps 31;

QY 2 ANISLKLFOKAIOKGLKLTALFTTSTAII---MLTGS-----GVLGAAVTVTADGAELAA 52
DB 858 ANVTLOADTNSNTGLKKRTLTLGNISVECNLSLTGANANIVGNLSIAEDSTFKGE----- 913
QY 53 GTNIGPGAGAFVAGSTLQYTGAFVTDADVSVALDINNFAAGLFSVTGDISLGSVVDTG 112
DB 914 -----ASDNLNITGTFT-----NNGTANI-NIKGVVVKLGDINNKG 947

QY 113 GANKLAVNIDDLTLTLTGTTGTAAYGANPALLFOGQAAAANNNTYALGNITLGG---AN 168
 Db 948 GLN-----ITTNASGTQKTLIINGNITNEKGLNIRN- IKADAEIQIGGNISOKE 995
 QY 169 AGLTIASDP-DVLGPITLAGNIDGGIITDNTDA-----AINGTIGNTPA 213
 Db 996 GNLTISSDKVNITNQITIKAGVEGGR--SDSSEAEANLTIQTKELKLAGDLNISGFNKA 1053
 QY 214 AOISGASTLSLGA-----VIKATTKLTNAAPVLTN-----ANAVLTGAVD 258
 Db 1054 ETAKNGSDLTIGNASGGNADAKKVTDFKDKSKISTDGHNVTLNSEVKTSSGSSNAGND 1113
 QY 259 NTFG-----GDDVGV-----LNLGALSQVTKNIGNTNSLAT--ISVGAGTATLGA 303
 Db 1114 NSTGLTISKADVTNNVTSHKTIINISAAAGNVTKEGTTINATTSVEVTAQNGTIKN 1173
 QY 304 VIKATTKLTNAASVLTTLNAVLTGAVDNTTGGDNVGVNLSGALSQVTKNIGNTNSLAT 363
 Db 1174 ITSQNVV-TAPENLVTTENAVI-----NATS---GTVNISTKTGDIKGGTSTS--GN 1221
 QY 364 INIGAVATLDCAVIKATTKLTDDASVLIFTPNVVVTGAIDNTGNANKGVVIFTGASTV 423
 Db 1222 VNITAGNTLKVSNITGQDVTADAGALTITAGSTISA---TTGNAN-----I 1267
 QY 424 TONIGNTAVLAEVSVGAGLLQIQG-----GVVKANALNTDNASVVTFTGDSV----- 472
 Db 1268 TTKTGDIKGVSSGSVTLVATGATLAVGNISGNTVITADSGKLTSTVSTINGTNSV 1327
 QY 473 -----TGSTGGTLPATVNIAG-----ITLRAGGSAAANNIDFGAASNLEFNGPAGKNYL 524
 Db 1328 TTSQSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKN--GAAT---LTAESGKLQT 1381
 QY 525 ICTIANGNATLINA-----AGTVIANDVSIQV-----AOINTONKFIYNK 571
 Db 1382 TGSSTSSNGQTLTAKDSSIAGNINAANVLTNTTGLTTTGDSKINATSGTL-TINAKD 1440
 QY 572 ADVDLIDQAISFKGAAS 589
 Db 1441 AKLD-----GAAS 1448

RESULT 4
 US-10-092-880-9
 ; Sequence 9, Application US/10092880
 ; Patent No. US20020164354A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J.
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/092,880
 ; CURRENT FILING DATE: 2002-03-08
 ; PRIOR APPLICATION NUMBER: 09/155,614
 ; PRIOR FILING DATE: 1998-09-30
 ; PRIOR APPLICATION NUMBER: 08/617,697
 ; PRIOR FILING DATE: 1996-04-01
 ; PRIOR APPLICATION NUMBER: PCT/US97/04707
 ; PRIOR FILING DATE: 1997-04-01
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 1599
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-092-880-9
 Query Match 9.9%; Score 299.5; DB 9; Length 1599;
 Best Local Similarity 22.8%; Pred. No. 2.3e-11;
 Matches 166; Conservative 98; Mismatches 247; Indels 217; Gaps 36;
 QY 19 TALFTTSTRAIML-----TSGSVLUG-AARTVTDAGELAGT-NIGPGAGA--FVAGSTLQ 70
 Db 670 TTLNVTSGSKFNLSIDSTGSGTGPSIRNAELNGITFNKATFNIAQSGSTANFSIKASIMP 729

QY 71 Y--TGAFVTVDADSVRA-----LDLNNFAAGL-----PSVTGDISLGSWDTG 112
 Db 730 FKSNNAYALFNEDISVSGGVSFNKLNASSNIQTPGVIIKSONENVSGGTLN--LKA 787
 QY 113 GANKLAVNIDDLTLTLTG-----TGT-----AAYGANPALLFOGQ-----AAANN 156
 Db 788 GSTETAFSIENDLNLNATGGNITIRQVEGTSRNVKNGVAACKNITFKGNITFFGOKATT 847
 QY 157 TALGNITLG-CANAGLTITASDPDVLGPITLAGNIDGGIITDNTDAAINGTIGN----- 209
 Db 848 EIKGNVTINKNTNATLRGANFAENKSPNLAGNVINNGNLT--TAGSINIAGNLTIVSKG 905
 QY 210 -----TNPAAOI-----SIGASTLSL--GGAVIKATTKLTNAAPVLTNAN-----AV 252
 Db 906 ANLOAITNVTNFAVAGSPDNNGASNISIRAGGAKFK---DINNTSLNITNDSITYRTI 961
 QY 253 LTGAVDNTTGGDDVGLN-----LNGALSQVTKNIG-----NTNSLATISVGA--- 295
 Db 962 IKGNISNKG--DLNIIDKSDAEIQIGNISQKEGNTLTISSDKVNITNQITIKAGVEGG 1019
 QY 296 -----GTATLGGAVIKATT-----GTATLGGAVIKATT- 309
 Db 1020 RSDSSEAEANLTIQTKELKLAGDLNISGFNKAETITAKNGSDLTIGNASGGNADAKKVT 1079
 QY 310 -----TKLTNAASVLTTLNAVLTGAVDNTTGGDNVGVNLSGALSQVTKNIGNTNSLATI 364
 Db 1080 DKVKDSKISTDGHNVTLNASEVKTSSGSSNAGNDNSTGLTISA--KDVTVN--NNVTSKTI 1136
 QY 365 NIGA---GVATLDGAVIKATTKL-----TDDASVLIFTPNVVVTGAIDN- 406
 Db 1137 NISAAAGNVTKEGTTINATTSVEVTAQNGTIKGNITSONVTVTATENLVTTENAVINA 1196
 QY 407 -----TGNANKGVVIFTGASTVDNIGNTAVLAE-----VSVGAG-LLQIQG 448
 Db 1197 TSGTVNISTKTGDIKGGIESTSGNVNITAS--GNTLKVSNITGQDVTVTADAGALTITAGS 1255
 QY 449 VVKANALNTDNASVVTFTGD--STVTSIGGTLPATVNIAGITLRAGGSALANNIDF 506
 Db 1256 TISAT-----TGNANITTKGDIKGVSSGSVTLVAT-----GATLAVGNI-- 1298
 QY 507 GAASLEFNGPAGKNYNLIGTIANGNATPLNINAG-----TVIANDVSIQT 553
 Db 1299 -SGNTVITITADSGKLTSTVGSTINGTNSVTSSQSDIEGTISGNTVNVNTASTGDLTIGN 1357
 QY 554 VAQINIQN 561
 Db 1358 SAKVEAKN 1365

RESULT 5
 US-10-184-644-83
 ; Sequence 83, Application US/10184644
 ; Publication No. US2003004930A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C227
 ; CURRENT APPLICATION NUMBER: US/10/184,644
 ; CURRENT FILING DATE: 2002-06-28
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612

```

; SEQ ID NO 83.
; LENGTH: 3127
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-184-644-83

```

	Query Match	9.7%	Score 292.5;	DB 9;	Length 3127;
	Best Local Similarity	26.2%;	Pred. No. 1.4e-10;		
	Matches 160;	Conservative 21;	Mismatches 307;	Indels 123;	Gaps 18;
QY	19	TALFTTSTAAILMTGSGVLGAARTVDAGAEAAAGTNIG-----PCAGAFVAGSTLQYT	72		
Db	1599	TAITTTCTTACCRAAGTCATGTGATGTGAAGTAGAATGAATGCACAGAGAAGTTTAT	1658		
QY	73	GATVTDAADVSRALDLNNFAAGLSFVSVDGISLVSVDTGGANKLAYNIDDLTLT----	128		
Db	1659	GAATATGGTGATGAGTTAGTAAAA--AGTGGCCATTATTGGCTTATTCTCGCTCTATAG	1716		
QY	129	LTCGTGAAYGANPALLFGGCAANNNTYALGNITLGGANAGLTIASDPDVLGPITILAGN	188		
Db	1717	TTGTGAAAATGAAGA---GTAAAAACAATTTGTTGACTATTTTAA-----AATTATA	1766		
QY	189	IDGGGIITDNDAINGT-----IGNTPAAQISIGASTLSLGGAIVIKATTTKLTNAA	241		
Db	1767	TTAGACCT--TAAGCTGTTTTAGCAAAGCATTAAGCAAATGTATGGCTGCCTTT--TGAA	1822		
QY	242	PVLTLLINAVNLGAVDNTTGGDDVGVLNUNG--ALSQVTCNGINTNSLATISVGAGTATL	300		
Db	1823	ATATTTGATG--TGTTGCCCTGGCAGGACTACTGCAAGAACATGGTTTATTTTAAAAATTA	1880		
QY	301	GGAVIRKA-----TTTKLTNAASVLTILTNAVLTGAVDN-----TTGGDNV	339		
Db	1881	TAAACAAGTCACCTAAATGCCACTTGCTCAAAAATCTTATAAGGTTTTACCCCTGTGATAC	1940		
QY	340	GVVNLSGALSQVNTGNINTNSLATINIGAGVATLDGAVIKATTTKLTFDDASVLIFTNPVV	399		
Db	1941	GGAAITTCACACAGGTAGGGAGTGTTTAGTGGAC--AATAGTGT-----	1981		
QY	400	VTCADINTGNANKGVVIFTCASVTVDNIGNTAVLAEVSVCAGLLQIOGGVVKANAINLTD	459		
Db	1982	-AGGTTATGGA-----TGGAGGTGTCGGTACTAATTTGAATAACGAGTAAATAATCTT	2033		
QY	460	NASV-----VFTEGDSTVTSIGGTGFATVN-----	486		
Db	2034	ACTTGGGTAGAGATGGCCCTTTGCCAACAAAGTGAACCTGTTTTTGTTGTTTTTAAACTCATG	2093		
QY	487	-----IGAGTTLRAGGSALAANNIDFGAASNLEPNPGAKYNULIGTYANGNNATLINAA	541		
Db	2094	AAGTATGGGTTTCAGTGGAAATGTTTCGAA--CTCTGAAGGATTTAGACAAGGTTTTGAAAA	2152		
QY	542	GTVIANDVSIIG	552		
Db	2153	GGATATCATCG	2163		

RESULT 6

```

US-10-092-880-4
; Sequence 4, Application US/10092880
; Patent No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; TITLE OF INVENTION: HAEMOPHILUS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-680-4

```

Query Match 9.5%; Score 286; DB 9; Length 1477;
Best Local Similarity 21.9%; Pred. NO. 1.5e-10;
Matches 173; Conservative 89; Mismatches 241; Indels 288; Gaps 36;

Qy	42	TVYADGAEIAAGTNI---GPGAGAFVAGS-----TLQYTGAFTVT-----78
Dd	597	IITTEGKDFRA-NNVSLNGTGKGLNIISSVNNLTHNLSGTINISGNITNOTTKRNTSYW655
Qy	79	----DADVSRALDLN---NF-----AAGL--FSVTCDIS-----104
Dd	656	QTSHDSHNWVSALNLETGANFTFIKYISSNKGLTTQYRSSAGVFNFGVMNSFLNKEG-715
Qy	105	-----LGSVDDTGGAN-----KLA-VNIDDG124
Dd	716	AKVNEFKLPENNNWTSKPLPIRFLANIATATGGSVFFDIYAHNSRGAEAKMSEINISG775
Qy	125	LILTLTG--TGTAAYCANFALL-----FQGQA--AANNYYTAL---GN161
Dd	776	ANPFLNSHVGRDDAFKINKDLTINATSNFSLURTKDDFYDGYARNAINSTYNISILGN835
Qy	162	ITLGGANAGLTIASD-----PD-----VLGPITLAG-NI189
Dd	836	VILGGONSSSIIGNITIEKANVYLEANNAPNOQNIRDRVIKLSLVNGSLUTGENA895
Qy	190	DGGGIITDNMTDAANGTIGNT-----NPAAQISICASTSLSGGAVIKATTTKLTNA240
Dd	896	DIKGNLTISATESATPKGKTDLTNITGNFTNNGTAENITQGVKLGNVTNDGDLNITHA955
Qy	241	APVLTILTANAVLTGAVDNTTGGDDV-----GVNLNGALSQVTCGTG-----NNNSL288
Dd	956	K-----RNORSIIIGDIIINKKGLSLTINSNDAEIQIGENISOEGNLTISSDKINITKQ1010
Qy	289	ATIISVG-----AGTATLGGAVIKATTTKLTNAASVLTLTNAVLTG-AVDNNTGGDNV339
Dd	1011	ITIKKGDGEDSSSDATSANLTIKTKEKLTEDLSISGFNKAEITAKDGRDLTIGNSND1070
Qy	340	G-----VVNLGALSQVTCGIG-----356
Dd	1071	GNSGAEAKTVTFENNVDKSKISADGHNVLTNSKVKTSSNGGRESNDTGLTITAKNVE1130
Qy	357	---NTNSLATINIGAG--VATLDGAVIKAT-----TTKLTDDASVLIFTPNPVVVTGAID405
Dd	1131	VNKDTISLKVNTIVASEKVTTTAGSTINATNGKASITTK-TGDISGTISGNTSVSATVD1189
Qy	406	NTGNANKGVPIFTGASTVTDNIGNTAVIAEVSVGAGLLIQIOGGVVVKANAILTNDASVWT465
Dd	1190	LTTKSGSKIEAKSGEANVTSATGTI-----GGVISGNTVNVTANAGDLT1233
Qy	466	FTGDSVTVIGSIGGETLFPATVNIAGITIPACGSILAANNIDFGAASNLEFFNGPACKKNVLI525
Dd	1234	VNGAEINATEGAAPTLTAT--GNLTITEAGSSITSTK---GOVDLLAQNG-----SIA1281
Qy	526	GTIANGNATLINAAAGTVIANDVSGIVNAIQIO-NNKIFVINAKADV-----DILD578
Dd	1282	GSI---NAANYTLNTGT-----LTVIAGSDIKATSGTLIVINAKOKLINGDASGDSTE1331
Qy	579	AQAISFKGAAS589
Dd	1332	VNAVNASGSGS1342

RESULT 7
US-09-886-468-21
; Sequence 21, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:

```

Query Match      9.1%; Score 276; DB 10; Length 871;
Best Local Similarity 23.1%; Pred. No. 3.6e-10;
Matches 147; Conservative 77; Mismatches 269; Indels 142; Gaps 25;

QY    30 MLTSGSVLGAARTVTADGAEIAAGTNI---CPGAGAFVAGSTLOQTGAFTVTDADVSVRA 86
       :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db     8 LLTSSALVFSLHPLMAANTDLSSSDNYENGSSGAATKETSASGT-----55

QY    87 LDLNFAAGLFVETGDLSGSV-----VDTGANKLAVNIDDL---TLTLGT 132
       :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db     56 -----TYTLSDVSITNVSAITPADKSCFNTGGALSF-VGADSHSLVLQTIALTHD 105

QY    133 GTAAYGANPALLFOG-----GQAANNYYTALGNITLGGANAGLTIASDPDVLPITLA 186
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     106 GAAINNTALISFGFSLLIDSAPATGTSGGKAICVNTTEGGTATFDN---ASVTLQ 162

QY    187 GNI---DGGGIITONTDAIINGTIGNTPAAQISIGASTLSIGGAVIKATTTKLITNAAPV 243
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     163 KNTSEKDGAAYSAYSIDLA-----KTITAALLQONTSTKNKGALCSTANTTVOGNSGV 216

QY    244 LTLN-----ANAVLTGAVDNT--TGG-----DDVGVLNLNGAL---275
        :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db     217 TFSSENTATDKGGGIYSKEKDSLTLDANTGVVTFKSNATAKTGGAWSDDDLALTGNQVLFQ 276

QY    276 -SQVTGNIGNTNS-----LATISVCAGCATLGGAVIKATTTKLITNAASVLILTN 323
        :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db     277 ENKTTGSAQAQNPEPGCGGAICCYLATATDKTGLAISQNMSETNTTANGGAIYATK 336

QY    324 AVLTGAVDNT--TGGDNVGVVNLGALSQVNTGNIGNTNSLATINIGAGVATLDGAVIKAT 381
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     337 CTLDG---NTTLTFDONTATACCGGAIYETETEDFSLKSGTGVTFSTNTAKTGALYSKG 393

QY    382 TTKLTDASVLIFNPVVVVTGAIDNTGN---ANKRGVLIIFTGASTVDNIG-NTAVLAEVS 437
        :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

Db 1230 SGDIGTSGGVEVKATESLTSQNSKIKATTGEANVTSATGTI-----1274
QY 446 QGGVVKANALNTDNASVVF-----TGDSTVTGSIIG-----TELFATVNI 487
Db 1275 -GGTISGNVNVNANAGDLTVNGAEINATECAATLTSSGKLTTAEASHITSAGQVNL 1333
QY 488 -----CAGITLRAGGSLA-----ANNIDFGA-----ASMLEFNGPAGKNYNLIG 526
Db 1334 SAQDSVAGSINAANVTLTGTTLTVKGSNINATSGTLVINAKDAELNGAALGNHTVV- 1392
QY 527 TIANGNNATLNTNAAGTVIANDVSIGTVAQINQNNKFFVINAKNADVLDILDAQAISPKG 586
Db 1393 -----NAT-NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----IISK 1429
QY 587 AASRLFLANVSLQMIELSLKKIYFVLLTV 616
Db 1430 GINTVLLKGVKID-----VKYIQPGIASV 1453

```

RESULT 9

```

US-10-184-644-39
; Sequence 39, Application US/10184644
; Publication No. US20030044930A1

```

GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin

```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227

```

```

; CURRENT APPLICATION NUMBER: US/10/184, 644
; CURRENT FILING DATE: 2002-06-28

```

```

; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612

```

```

; SEQ ID NO 39

```

```

; LENGTH: 1819

```

```

; TYPE: DNA

```

```

; ORGANISM: Homo Sapien

```

```

US-10-184-644-39

```

```

Query Match 9.0%; Score 271.5; DB 9; Length 1819;

```

```

Best Local Similarity 27.9%; Pred. No. 1.6e-09;

```

```

Matches 147; Conservative 16; Mismatches 276; Indels 87; Gaps 19;

```

```

QY 19 TALFTTTSTAALMLTSGVGLGAARTVTADGAELAAAGTNIIGPGAGFVAGSTLQYTGFTVT 78

```

```

Db 1345 TGTATTTCTAGAAAT-AGTTATGTCTTAGGAAATGT-----GCTTTAATTTTGACTTT 1397

```

```

QY 79 DADVSVALDLNNFAAGLFSVTGDISLG--SVVDTGANKLAVN-IDDGLTLT-LTGTGT 134

```

```

Db 1398 TA-----CAGGTAAAGTGCAGAGGAGAGTGTTCATGAAATGTTCTAATGTAT 1446

```

```

QY 135 AAYG-----ANPALLFQGGOA-AANNNTYALGN-ITLGGANAGLTI 173

```

```

Db 1447 AATAACATTTACCTTCACGCTCCATCAGATGGAACGAGTTTCAGTAATCAGGAATAT 1506

```

```

QY 174 ASDPDVLGPTITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKAT 233

```

```

Db 1507 AT-----CTATATGATCTTGATATTTTATAATAAATTTGAA-----GTCTAAAGACTGCA 1559

```

```

QY 234 TTKLTNAAPVLTLTNANAVLTGAVDNTGG--DDVGVLNLNGALSQVTVGNTGNTSLATI 291

```

```

Db 1560 TTTTAAACAAGATTAGTATTAATGCGTGGCCACGTAGCAAAAGATAT-----1609

```

```

QY 292 SVGAGTATLGGAVIKATTTKLTNAAASVLTLTNAVLTGAVDNTT-----GGDNVGVNLSGA 347
Db 1610 TTGATTAATCTTAAAAATGTTAAATACCGTTTTCATGAAATTTCTCAGTATTTGTAACAGC 1669
QY 348 LSOVTGNTGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIFTPVNVVVTGAIDNT 407
Db 1670 AACTTGTCAAAACCTAA--GCATATTTGAATATGATCTCCCATATATTTGAAATTTGAAATC 1726
QY 408 GNANKGV-----VIFTGASTVTDNIGNTAVLAEVSVGAGLLQIOGGVVKANALNTDNASV 463
Db 1727 GTATTGTGTGCTCTGTATATCTGTTAAAAAATTTAAA-----GGACAGAAACCT---TT 1778
QY 464 VTFTGDSVTGTSIGSTELFATVNVIGAGITLRAGGSLAANNIDFGAA 509
Db 1779 CTTTGTGTATGCATGT-----TTGAATTAAGAAAGTAATGGAA 1818

```

RESULT 10

```

US-10-184-644-379

```

```

; Sequence 379, Application US/10184644

```

```

; Publication No. US20030044930A1

```

GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin

```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227

```

```

; CURRENT APPLICATION NUMBER: US/10/184, 644
; CURRENT FILING DATE: 2002-06-28

```

```

; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612

```

```

; SEQ ID NO 379

```

```

; LENGTH: 2340

```

```

; TYPE: DNA

```

```

; ORGANISM: Homo Sapien

```

```

US-10-184-644-379

```

```

Query Match 8.5%; Score 258; DB 9; Length 2340;

```

```

Best Local Similarity 24.1%; Pred. No. 1.6e-08;

```

```

Matches 147; Conservative 18; Mismatches 320; Indels 124; Gaps 16;

```

```

QY 16 GLKTALETTTSTAALMLTSGVGLGAARTVTA-----DGAELAAAGTNIIGPG-----59

```

```

Db 885 GATTATATCTCATGCGCAAAAGCTTTATAATTGCTGGCTTAGGACAGCAATACTTTA 944

```

```

QY 60 --AGAFVAGSTLQYTGFTVTDADVSVRLDLNNFAGLFSVTGDISLGSVVDTGANKL 117

```

```

Db 945 CAATAAAGCTCTACACATTTTC-----AAGGAGTATGCTGGATTCATGGA---990

```

```

QY 118 AVNIDDDLTLTGTGTAAAYGANPALLFQGGQAAANNNTYALGNITLGGANAGLTIASDP 177

```

```

Db 991 ACTCTAATCTGTACATAAAAAATTTTAAAGTATTTGCTTTTCAGGCAAGTCGTCTTC 1050

```

```

QY 178 DVLG-----PITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSLG 226

```

```

Db 1051 AATGCTGTACTATGCTCTCTTAAGAGAAATTTGGTAACCTTGTGTATGTTGGTAAGCATAG 1110

```

```

QY 227 GAVIKATTTKLTNAAPVLTLTNANAVLTGA---VDNTTGGDDVG---VLNLNGALSQVT 279

```

```

Db 1111 GTGAGTTTTGTATAAATCTTTTGTGTTGAGATCAAGCTGAAATGAAACACTGAAAAAC 1170

```

```

QY 280 GNTGNTNSLATISVGAGTATLGGAVIKATTTKLTNAAASVLTLTNAVLTGAVDNTTGGDNV 339

```

```

Db 1171 ATGGAATTCATT-----TCTATAACACATTTATTTAAGTATATAACACGTTTTTTTGACA 1224

```

QY 340 GVVNLGSLQSVT-----GNIGNTSLATINIGAGVATLDG 375
 Db 1225 AGTGAAGAAITGTTAATCATCTGCTCATTTGCTCAATAGATGTAACCTGTAGACTACG 1284
 QY 376 AVIKATTTKTLDDASVLIETPNVVTGAIDNTG--NANKGVVIFTGASTVTDNIGNTAVL 433
 Db 1285 G-CTATTTG-----AAAAAATGCTTATTGCTACTATATTTGTTATTCGA 1329
 QY 434 AEVSVCAGLLQIQGGVVKANINLTDNASVVTFTGDSVTGSGTGTELFATVNVIGAGITL 493
 Db 1330 ATATCAG-----CAGAGAAAGGAAATATATGTTGAAATAAATGTTTGAATC 1379
 QY 494 RAGGSLAANNIDFGAASNLFECPACKNYNLIG-----TIANGNNATLINAAGTV 544
 Db 1380 ATGACCCA--AAGAAAT-----GTATTGATTCACACTATCCITCAGAACTGAAGGTT 1431
 QY 545 IANDVSIGT 553
 Db 1432 AATTATTGT 1440

RESULT 11
 US-09-996-634-131
 ; Sequence 131, Application US/09996634
 ; Patent No. US20020172684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; FILE REFERENCE: 61260
 ; CURRENT APPLICATION NUMBER: US/09/996,634
 ; PRIOR APPLICATION DATE: 2001-11-28
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08/990,823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-996-634-131

Query Match 8.5%; Score 256.5; DB 9; Length 943;
 Best Local Similarity 24.6%; Pred. No. 6.9e-09;
 Matches 171; Conservative 79; Mismatches 250; Indels 195; Gaps 35;

QY 10 QKAIQKGLKLTALTTSTAA-----IMLTGSGVLGA----- 39
 Db 66 QAATQRLAAAFEAALAAATVHPAIIISANRALFVSLVSNLLGQNPATAATEAAYEQMW 125
 QY 40 ARTVTA-----DGAELAAGTINIG-----PCAGAFVAGSTLQYT-----GFTVTD 79
 Db 126 AQDAAMFCYHAGASAASVLTFFGQALPTVAGGGALVSAASAAQVTRVFRNLGLANVGE 185
 QY 80 ADVSRALDLNINFAAGLFSV--TGDISLGSVWDTGGANKLAVNIDDLGLTLTGTGTAAYG 138
 Db 186 GNVGNG--NVGNFNLGSANIGNGSGNI---GSSNIGFNGVGPGLTAALNNIGFGNTG 240
 QY 139 ANPALLFOGQQAANNYYTALGNITLGGANAGLTIASDPDVLGPITLAGNIDGGGIITDN 198
 Db 241 SNN---IGFGNTGSSN--IGFNGTGDNRGIGLT---GSGLLG-----FGLNSG---TCN 285
 QY 199 TDAANGT-----IGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTITNANAVLT 254
 Db 286 IGLFNSGTGNGVGNIGS-----GTGNWNGIG-----NSGNSYNT 317

QY 255 GAVDNTTGGDDVGVNLNGLSALSOVTGNTGNNTSLATISVGAGTATLGGAVIKATTTKLTN 314
 Db 318 GF--GNSGDANTGFFN--SGIANTGVCNAGNYN---TGSYNPNSNTGGFNMGOYNTGYLN 371
 QY 315 AASVLT-----LFAVLTGAVDN-----TTGDNVGV 341
 Db 372 SCNYNTGLANSNVNTGAFITGNFNGFLWRGDHQLIFGSPGFFNSTAPSSGFFNSGA 431
 QY 342 VNLGSLASQVTVGNIG--NTNSLATINIG--AGVATLDGAVIKATTTKLTDDASVLIFTN 396
 Db 432 GSASGFLNSGANNSSGFFNSSGAIGNSGLANAGV--LVSGVINSNTVSGLFNMSLVAIT 490
 QY 397 PVVVTG-----AIDNTGNANKGVVIFTGASTVTDNIGNTAFLAELVSVGAG 441
 Db 491 PALISGFFNTGSMGSGFFGPPVFNGLANRGVNVNIGNA---NIGNVNLGSGNVGDF 546
 QY 442 LLQIQGGVVKANAINLTDNASVVTFTGDSVTGSGIG---GTELFATVNIQA---GITLR 494
 Db 547 NILGSGNL--GSQLILGSGNVGSPNIGS--GNIGFNVGSGSLGNYNIGSGNLGIYNI 600
 QY 495 AGSLAANNIDFGAASNLFEFN--GPACKNYNLIGTIANGNNATLNI-----NAAGTVIA 546
 Db 601 GFENVGDYNVGFGNAG--DFNOGFANTGNNNIGFANTGNN---NIGTGLSGDNOQGFNIA 655
 QY 547 NDVSICT--VAQINIONNKIEFVINAKNADVDILDA 579
 Db 656 SCHNSGTGNSGLFNSCTNNVGFNAGTGNVGIANS 690

RESULT 12
 US-09-997-181-131
 ; Sequence 131, Application US/09997181
 ; Publication No. US20030049269A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; FILE REFERENCE: 61257
 ; CURRENT APPLICATION NUMBER: US/09/997,181
 ; CURRENT FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08/990,823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-997-181-131

Query Match 8.5%; Score 256.5; DB 9; Length 943;
 Best Local Similarity 24.6%; Pred. No. 6.9e-09;
 Matches 171; Conservative 79; Mismatches 250; Indels 195; Gaps 35;

QY 10 QKAIQKGLKLTALTTSTAA-----IMLTGSGVLGA----- 39
 Db 66 QAATQRLAAAFEAALAAATVHPAIIISANRALFVSLVSNLLGQNPATAATEAAYEQMW 125
 QY 40 ARTVTA-----DGAELAAGTINIG-----PCAGAFVAGSTLQYT-----GFTVTD 79
 Db 126 AQDAAMFCYHAGASAASVLTFFGQALPTVAGGGALVSAASAAQVTRVFRNLGLANVGE 185
 QY 80 ADVSRALDLNINFAAGLFSV--TGDISLGSVWDTGGANKLAVNIDDLGLTLTGTGTAAYG 138
 Db 186 GNVGNG--NVGNFNLGSANIGNGSGNI---GSSNIGFNGVGPGLTAALNNIGFGNTG 240
 QY 139 ANPALLFOGQQAANNYYTALGNITLGGANAGLTIASDPDVLGPITLAGNIDGGGIITDN 198

```
Db 241 SNN---IGFGTGSNN---IGFGTGDGNGRGIGLT---GSGLLG-----FGGLNSG---TGN 285
Qy 199 TDAINGT---IGNTNPAAGISGASTLSLGGAVIKATTTKLTNAAPVLTITNANAVLT 254
Db 286 IGLFSGTGNVGNIGNS-----GTGNWIG-----NSGNSYNT 317
Qy 255 GAVDNTTGGDDVGVNLNGALSQVNTGNIGNTNSLATISVGAGTATLGGAVIKATTTKLTN 314
Db 318 GF---GNSGDANTGFEN---SGIANTGVGNAGYN---TGSYNPGNSNTGGFNMGOYNTGYLN 371
Qy 315 AASVLT-----LTNAVLTAVDN-----TTGDDNVGV 341
Db 372 SGNTNTGLANSNTNTGAFITGNFNGFLWRGDHGLIFGSPGFNSTSAPSSGFFNSGA 431
Qy 342 VNLGALSQVNTGNIG---NTNSLATINIG---AGVATLDGAVIKATTTKLTDDASVLIPTN 396
Db 432 GSASGFLNSGANNSSGFFNSGSAIGNSGLANAGV---LVSQVINSNTVSGFLNMSLVAIT 490
Qy 397 PVVVTG-----AIDNTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAG 441
Db 491 PALISGFFNTGSNMGFFGPPVFNGLANRGVYNILGNA-----NIGNYILGSGNVGDF 546
Qy 442 LLIQGGVVKANALNTDNASVFTTGDSTVTGSG-----CTELFATVNICA---GITLR 494
Db 547 NILGSGNL---GSONILGSGNVGSFNIGS---GNIGFNVGSGSLGNYNIGSGLGIYNI 600
Qy 495 AGGSLAANNIDFGAASNLEFN---GPAGKYNILGTTIANGNNATLNI-----NAAGTVIA 546
Db 601 GFGNVGDYNVGFGNAG---DFNOGFANTGNNGNIGFANTGNN---NIGIGLSDNQOQFNIA 655
```

RESULT 13

```
US-09-997-182-131
; Sequence 131, Application US/09997182
; Publication No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131
```

```
Query Match 8.5%; Score 256.5; DB 9; Length 943;
Best Local Similarity 24.6%; Pred. No. 6.9e-09;
Matches 171; Conservative 79; Mismatches 250; Indels 195; Gaps 35;
```

```
Qy 10 QKAIQKGLKTLFTSTAA-----TMLTSGSVLGA----- 39
Db 66 QAATQTRLAFAAFAALAAVHPALISANRALFVSLVSNLLGQNAFAIAATEAAYEOMW 125
Qy 40 ARTVTA-----DGAELAGTNTG-----PGAGAFVAGSTLQYT-----GAFVTVD 79
```

```
Db 126 AQDVAAAMEGYIAGASAAVSALTPEQQAALPTVAGGALVSAAAAQVTTTRFNLGLANVGE 185
Qy 80 ADVSRALDLNNAAGLFSV---TGDISLGSVVDTGANKLAVNIDDLTLTITGTGTAAYG 138
Db 186 GNVGNG---NVGNFNLGSANIGNIGSGNI---GSSNIGFNGVGPGLTAALNNIGFGWTG 240
Qy 139 ANPALLFOGCGAANNTYTALGNTITLGGANAGLTIASDDVLPITTLAGNTDGGIITDN 198
Db 241 SNN---IGFGTGSNN---IGFGTGDGNGRGIGLT---GSGLLG-----FGGLNSG---TGN 285
Qy 199 TDAINGT---IGNTNPAAGISGASTLSLGGAVIKATTTKLTNAAPVLTITNANAVLT 254
Db 286 IGLFSGTGNVGNIGNS-----GTGNWIG-----NSGNSYNT 317
Qy 255 GAVDNTTGGDDVGVNLNGALSQVNTGNIGNTNSLATISVGAGTATLGGAVIKATTTKLTN 314
Db 318 GF---GNSGDANTGFEN---SGIANTGVGNAGYN---TGSYNPGNSNTGGFNMGOYNTGYLN 371
Qy 315 AASVLT-----LTNAVLTAVDN-----TTGDDNVGV 341
Db 372 SGNTNTGLANSNTNTGAFITGNFNGFLWRGDHGLIFGSPGFNSTSAPSSGFFNSGA 431
Qy 342 VNLGALSQVNTGNIG---NTNSLATINIG---AGVATLDGAVIKATTTKLTDDASVLIPTN 396
Db 432 GSASGFLNSGANNSSGFFNSGSAIGNSGLANAGV---LVSQVINSNTVSGFLNMSLVAIT 490
Qy 397 PVVVTG-----AIDNTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAG 441
Db 491 PALISGFFNTGSNMGFFGPPVFNGLANRGVYNILGNA-----NIGNYILGSGNVGDF 546
Qy 442 LLIQGGVVKANALNTDNASVFTTGDSTVTGSG-----CTELFATVNICA---GITLR 494
Db 547 NILGSGNL---GSONILGSGNVGSFNIGS---GNIGFNVGSGSLGNYNIGSGLGIYNI 600
Qy 495 AGGSLAANNIDFGAASNLEFN---GPAGKYNILGTTIANGNNATLNI-----NAAGTVIA 546
Db 601 GFGNVGDYNVGFGNAG---DFNOGFANTGNNGNIGFANTGNN---NIGIGLSDNQOQFNIA 655
```

RESULT 14

```
US-10-184-644-5
; Sequence 5, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```

```
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 5
; LENGTH: 1487
; TYPE: DNA
; ORGANISM: Homo Sapien
```

```
US-10-184-644-5
```

```
Query Match 8.4%; Score 254; DB 9; Length 1487;
```



```

Query Match      8.4%; Score 253.5; DB 9; Length 2397;
Best Local Similarity 25.4%; Pred. No. 3.le-08;
Matches 149; Conservative 15; Mismatches 292; Indels 131; Gaps 20;

QY      19  TALFTTTAAIMLTGSOVLGAARTVTADGAEALAGTWNIGFCAGAFVAGSTLQYTGAFVTT 78
Db      857  TAAATTTTATAGTTATCTTAAATATATATGATTTTGATAAAACACAGAAGT---TGATCAT 912

QY      79  DADVSVALDNNPAAGLFSVTGDISGVSVDYTGANKLAVNIDDLGLTLTLTGTTGTAAYG 138
Db      913  -----TTTGTTGGTTTGAAGTGAACGTGTGACCTTTTGTGAATATTG 953

QY      139  ANPALLFQGGGAANNNTYTA-----LGNITLGCANAGLTIASDPDVLGPITLAGNI 189
Db      954  C-----AGGTTTCAGTCTAGATTGTCATTAAATGAAGAGCTCTACATTCAGAACATATA 1007

QY      190  DG-----GGIIT-----DNTDAAINGTIGNTPNPAQISIGASTLSLGGAVIKATTTKLTN 239
Db      1008  AGCACTAGGTATACAAGTTTGAATATATGATTTAAGCACAGTATCATG-----TTTAAAT 1062

QY      240  AAPVLTLTNNAVLTGAVDNLTGGDDV-----GVNLNGALSOVTGN----- 281
Db      1063  AGTTCTCTAAATTTTGAATAAATCGTCCACGAATAAGATTATGATATATGTTGTTAATA 1122

QY      282  -----IGNTNSLATISVGAGTATLGGAVIKATTTKLTNNAASVLTLTNAVLTGAVDNNTG 335
Db      1123  ATACCTATTTCAGTCTGAGTTTGGAAA--ATTTCACATTTCCCAAGTATTCGCAATTATG 1180

QY      336  GDNVGVVNLGALSQVLTGNNTNSLATINITAGV-----ATLDGAVIKATTT 383
Db      1181  AGGTATTTAACAAGATTA-----TTTTAGACAAAAATATTTCTCATTTGATATAATT 1232

QY      384  KLTDASVLIFTPNPVVTGAIDNTGNANKGVVITFGASTVTDNIGNTAVLAEVSVGAGLL 443
Db      1233  TTT---CTCTGTTTCACTGT--GTGAASAAAAGAGATATTTCCCATATA---AATGGGAA 1283

QY      444  QIQGVVVKANAINLTDNASVVTFGDSVTGTSIGGT--ELFATVNIAG-----ITILRAG 496
Db      1284  GTTTGGCCCATGTCTCAAGAAA--TGCTGATTTTCAGTGCAATTTTCGTGGTCTTTTAGAG 1342

QY      497  GSLAANNIDFGAASNLBFGNPGAGKNYLNIGTIANGNNATLINAAGT 543
Db      1343  GTATATTCCAAATTT-----TCCTTGATTTTTTAGGT 1374

```

Search completed: April 14, 2003, 16:12:05
Job time : 33 secs

```

RESULT 15
US-10-184-644-29
; Sequence 29, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 29
; LENGTH: 2397
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-29

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:08:59 ; Search time 150 Seconds

(without alignments)
2660.599 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKFQKAIQKGLKTA.....MIELSLKLIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3022	100.0	619	22	US-09-800-065-2
2	1494.5	49.5	2249	4	US-08-029-329-2
3	350	11.6	4327	27	US-60-360-039-10178
4	333.5	11.0	1026	3	US-07-895-367-7
5	333.5	11.0	1026	10	US-08-614-377-7
6	331	11.0	1026	5	US-08-194-290A-7

7	331	11.0	1026	17	US-09-379-931-7	Sequence 7, Appli
8	331	11.0	1026	21	US-09-743-731-5	Sequence 5, Appli
9	323.5	10.7	1242	21	US-09-739-449-13178	Sequence 13178, A
10	323.5	10.7	1242	22	US-09-803-110-13178	Sequence 13178, A
11	322	10.7	1180	15	US-09-167-568-61	Sequence 61, Appli
12	322	10.7	1180	15	US-09-167-568A-61	Sequence 61, Appli
13	322	10.7	1188	15	US-09-167-568-59	Sequence 59, Appli
14	322	10.7	1188	15	US-09-167-568A-59	Sequence 59, Appli
15	316.5	10.5	915	15	US-09-167-568-35	Sequence 35, Appli
16	316.5	10.5	915	15	US-09-167-568A-35	Sequence 35, Appli
17	316.5	10.5	1222	15	US-09-167-568-37	Sequence 37, Appli
18	316.5	10.5	1222	15	US-09-167-568A-37	Sequence 37, Appli
19	316.5	10.5	1228	15	US-09-167-568-34	Sequence 34, Appli
20	316.5	10.5	1228	15	US-09-167-568A-34	Sequence 34, Appli
21	303.5	10.0	1600	15	US-09-155-614-10	Sequence 10, Appli
22	303.5	10.0	1600	24	US-10-092-880-10	Sequence 10, Appli
23	301	10.0	1215	27	US-60-360-039-10073	Sequence 10073, A
24	299.5	9.9	1599	15	US-09-155-614-9	Sequence 9, Appli
25	299.5	9.9	1599	24	US-10-092-880-9	Sequence 9, Appli
26	299.5	9.9	2732	26	US-10-238-075-1119	Sequence 1119, Ap
27	298.5	9.9	2834	24	US-10-085-959-252	Sequence 252, App
28	296	9.8	1060	21	US-09-791-537-65965	Sequence 65965, A
29	296	9.8	1060	21	US-09-791-537-65966	Sequence 65966, A
30	296	9.8	1060	21	US-09-791-537-65967	Sequence 65967, A
31	294	9.7	1060	21	US-09-791-537-65971	Sequence 65971, A
32	293.5	9.7	969	15	US-09-167-568-32	Sequence 32, Appli
33	293.5	9.7	969	15	US-09-167-568A-32	Sequence 32, Appli
34	293.5	9.7	975	15	US-09-167-568-30	Sequence 30, Appli
35	293.5	9.7	975	15	US-09-167-568A-30	Sequence 30, Appli
36	293	9.7	1060	21	US-09-791-537-65972	Sequence 65972, A
37	293	9.7	1465	27	US-60-360-039-13955	Sequence 13955, A
38	292.5	9.7	1054	21	US-09-791-537-103009	Sequence 103009,
39	292.5	9.7	3127	25	US-10-176-912-83	Sequence 83, Appli
40	292.5	9.7	3127	25	US-10-179-524-83	Sequence 83, Appli
41	292.5	9.7	3127	25	US-10-184-634-83	Sequence 83, Appli
42	292.5	9.7	3127	25	US-10-184-644-83	Sequence 83, Appli
43	292	9.7	1612	3	US-07-742-128-2	Sequence 2, Appli
44	289	9.6	1220	15	US-09-167-568-28	Sequence 28, Appli
45	289	9.6	1220	15	US-09-167-568A-28	Sequence 28, Appli

ALIGNMENTS

RESULT 1
US-09-800-065-2
; Sequence 2, Application US/098000065
; GENERAL INFORMATION:
; APPLICANT: Bouyer, Donald H
; APPLICANT: Crocquet-Valdes, Patricia
; APPLICANT: Stenos, John
; APPLICANT: Walker, David H
; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein
; FILE REFERENCE: 026.00121
; CURRENT APPLICATION NUMBER: US/09/800,065
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: US 60/187,323
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Rickettsia felis
US-09-800-065-2

Query Match 100.0%; Score 3022; DB 22; Length 619;
Best Local Similarity 100.0%; Pred. No. 5.2e-247;
Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANISLKFQKAIQKGLKLTFTTAAIMLTGSGVLGAARTVTADGAELAGTNGPGA 60
|||||
DB 1 MANISLKFQKAIQKGLKLTFTTAAIMLTGSGVLGAARTVTADGAELAGTNGPGA 60
|||||

```

QY 61 GAFVAGSTLOQTGFTVTDADVSVRALDNNFAAGLFSVTDGDISLGSVVDGAGANKLAVN 120
DB 61 GAFVAGSTLOQTGFTVTDADVSVRALDNNFAAGLFSVTDGDISLGSVVDGAGANKLAVN 120
QY 121 IDGGLTTLCTGTAAGANPALLFOGQAAANNNTYALGNITLGGANAGLTIASDPPVL 180
DB 121 IDGGLTTLCTGTAAGANPALLFOGQAAANNNTYALGNITLGGANAGLTIASDPPVL 180
QY 181 GPITLGNIDGGIITDNTDAINGTNTNPAAQISIGASTLSLGGAVIKATTTKLTNA 240
DB 181 GPITLGNIDGGIITDNTDAINGTNTNPAAQISIGASTLSLGGAVIKATTTKLTNA 240
QY 241 APVLTNLNANAVLTCGDDVGVNLNGALSOQVTGNIGNTNSLATISVAGGATL 300
DB 241 APVLTNLNANAVLTCGDDVGVNLNGALSOQVTGNIGNTNSLATISVAGGATL 300
QY 301 GGAVIKATTTKLTNAASVLTLTNAVLTCGDDVGVNLNGALSOQVTGNIGNTNS 360
DB 301 GGAVIKATTTKLTNAASVLTLTNAVLTCGDDVGVNLNGALSOQVTGNIGNTNS 360
QY 361 LATINIGAGVATLDGAVIKATTTKLTDDASVLIETNPVVVTGAIDNTGNANKGVVIFGA 420
DB 361 LATINIGAGVATLDGAVIKATTTKLTDDASVLIETNPVVVTGAIDNTGNANKGVVIFGA 420
QY 421 STVTDNTGNTAVLAEVSVGAGLLQIQGVVYKANINLTNDNASVVTFTGDSVTGSGIGTE 480
DB 421 STVTDNTGNTAVLAEVSVGAGLLQIQGVVYKANINLTNDNASVVTFTGDSVTGSGIGTE 480
QY 481 LFAVNTGAGITLRAAGSLAANNIDFGAASNLFEFGPAGKNYNLIGTANGNATLINA 540
DB 481 LFAVNTGAGITLRAAGSLAANNIDFGAASNLFEFGPAGKNYNLIGTANGNATLINA 540
QY 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVSLOM 600
DB 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVSLOM 600
QY 601 IELSLKIIIPVLTVVVS 619
DB 601 IELSLKIIIPVLTVVVS 619

```

RESULT 2

US-08-029-329-2

Sequence 2, Application US/08029329

GENERAL INFORMATION:

APPLICANT: SUMNER W., JOHN

APPLICANT: ANDERSON E., BURT

APPLICANT: PELLET E., PHILIP

APPLICANT: SANCHEZ-MARTINEZ, DEMETRIO

TITLE OF INVENTION: "BACULOVIRUS EXPRESSED 190 kDa ANTIGENS

OF RICKETTSIA RICKETTSII AND RELATED METHODS"

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: NEEDLE & ROSENBERG, P.C.

STREET: 133 Carnegie Way, Suite 400

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/029,329

FILING DATE: 19930309

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SPRATT, GWENDOLYN D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 1414, 049

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770

TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2249 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-029-329-2

Query Match 49.58; Score 1494.5; DB 4; Length 2249;

Best Local Similarity 51.18; Pred. No. 1.1e-116;

Matches 364; Conservative 58; Mismatches 123; Indels 167; Gaps 18;

```

QY 20 ALETTSTAAILTGS-----GVL---GAARTVTAD--GAELAAGTNIQGP-----A 60
DB 613 SVLTLTANAVLTGAIDNTTGGDNVGVNLNGALSOQVTGDISLGNSTNSLATISVAGATATLG 672
QY 61 GAFVAGSTLOQTGFTV---TDADVSVRALD--LNNEAAGLFSVTDGDISLGSVVDGAGANK 116
DB 673 GAVIKATTTKLTNAVASAVKFTNPVVVTGAIDSTGNANNGIVTFTGNSVTVG--DIGNTNA 730
QY 117 LA-VNIDDLGLTLTCTGTAAYGANPALLFOGQAAANNNTYALGN-----ITLGGANAGL 171
DB 731 LATVNV-----GAGTATLGG-----AVIKATTTKLTNAASVLTLTNANAVL 771
QY 172 TIASD---PDVLGPITLAGNIDGGIITDNTDAINGTNTNPAAQISIGASTLSLGG 227
DB 772 TGAIDNTTGGDNVGVNLNGAL-----SOQVTGDISLGNSTNSLATISVAGATATLGG 820
QY 228 AVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDVGVLNGLALSOQVTGNIGNTNS 287
DB 821 AVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDVGVLNGLALSOQVTGDISLGN 880
QY 288 LATISVAGATATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTTGGDNVGVNLS 345
DB 881 LATISVAGATATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLN 940
QY 346 GALSQVTGNIGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIETNPVVVTGAID 405
DB 941 GALSQVTGDISLGNSTNSLATISVAGATATLGGAVIKATTTKLTDAASAVKFTNPVVVTGA 1000
QY 406 NTGNANKGVVIFTGASTVTDNIGNTAVLAEVSVGAGLLQIQGVVYKANINLTNDNAS--- 462
DB 1001 NTGNANKGVVIFTGASTVTDNIGNTAVLAEVSVGAGLLQIQGVVYKANINLTNDNASAVT 1060
QY 463 ----- 462
DB 1061 FTNPVVVTGAIDNTGNANNGIVTFTGNSVTVGNVGNLTNALATVNVGAGLLQVQGVVYKAN 1120
QY 463 -----VVTFTGDSVTVGSGISIGTELFATVNIAG 490
DB 1121 TINTDNASAVTFTNPVVVTGAIDNTGNANNGIVTFTGNSVTVGDISLGNLTNALATVNVGAG 1180
QY 491 ITRAGSLAANNIDFGAASNLFEFGP---AGK--NYNLIGTANGNATININAGTAVI 545
DB 1181 ITRAGSLAANNIDFGAASNLFEFGP---AGK--NYNLIGTANGNATININAGTAVI 545
QY 546 ANDVSGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVS 597
DB 1240 ASHLLTIGTVAEINAGNLTIDASVGVDTILNAQININFRARDSVLSNLT 1291

```

RESULT 3

US-60-360-039-10178

Sequence 10178, Application US/60360039

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Chen, Xianfeng

APPLICANT: Goldman, Barry S.

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE REFERENCE: 38-10(52052)A
;; CURRENT APPLICATION NUMBER: US/60/360,039
;; CURRENT FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 10178
;; LENGTH: 4327
;; TYPE: PRT
;; ORGANISM: magnetite-containing magnetic coccus
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (1)..(4327)
;; OTHER INFORMATION: unsure at all xaa locations
US-60-360-039-10178

Query Match 11.6%; Score 350; DB 27; Length 4327;
Best Local Similarity 25.6%; Pred. No. 3.3e-19;
Matches 169; Conservative 90; Mismatches 236; Indels 166; Gaps 31;
QY 16 GLK--TALFTTSTAALMLTSGVLGAARTV-----TADGAELAACTNIGPGAGAFV 64
DB 3061 GLKDGAAALAAANTS-YITGSVSVYDVAQMAQLSAMDTATGILTYGAGLKDSAAALV 3119
QY 65 AGSTLQYTGAFVTDADVS-----VRALD-----LNNFAAGLFSVTGDISLGS---VVDPT 111
DB 3120 ANTNSYVTGAVTVTDAATTAQLGAIDQDTTCTVNYSLAGIKDVTNSNITIDSGNYVANA 3179
QY 112 GGANKLAVNIDGLILTLTGTAAYGANPALLFQCGQAAA-----NNYTTALGNTTLGG 166
DB 3180 GGA---TITVNDGIANLITDAGTVGTGRNVTVDAAQMAQLSQIDNYTTGALKYVTIKD 3236
QY 167 ANAGLIAIDPDVLGPIITLGNIDGGIITDNTDAAINGT--IGNTNPAQISIGASTLS 224
DB 3237 AVAALVA-----NTNSVTSYAVSVYDAASMAQLSAIDQD 3272
QY 225 LGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTGGDDGVNLNGLALSOVTGNIGN 284
DB 3273 TTGTL---TYTKLTDAVANL-VTNTNSYVTGSV-NVTVSDIATISQLSSIDANTTGSVY 3327
QY 285 T---NSLATISVCACT--ATLGGAVIKATT-TKLT-----NAAS 317
DB 3328 TQIGDAAATLATNAGNYVRKATHVTVTDAATIAQLFTIDGDNNTGSLVYTAGGVKDSAA 3387
QY 318 VLTLTNAVLTGAVDNTTGGDNVGNLNGALSOVTGNIG-----NTNSLAT 363
DB 3388 LVVNTNSYVTGAV-NVSVTDVTSIAQLSAVDEYTTGTLTYGAGVKDSVANLLVNTNSYV 3446
QY 364 -----INIGAGVATLDGAVIKATT-----TKLTDDASVLIFTNPVVVTGAI-----DNT 407
DB 3447 GSYAVSITDVASMANL-SAIQDFTTGTGLNVTKLSDTVSALVANTNSYVTGSVNVITDNA 3505
QY 408 GNANKGVWIFTGASTVTDNIGNITAVLAESVAGLLOIOGGVVKANAIN--LTDNASVYT 465
DB 3506 SWAN-----MSAIDQNTGTLTYTKLSDDTAALAAANTNSYVTGSVNVTVTDAATVAQ 3557
QY 466 FTG-DSTVTSIGGTELFATVNIAGITLRAGGSLAANNIDFGAASNLFEFGPAGKNY--522
DB 3558 LITVDAATGTI---KVASVDSGSI-----SSNFAYVDGLGVSYVN 3597
QY 523 ---NLIG-TIANGNATLINNAGTVIANDVSGTVAQINQNNKIFVINAKNADVIDLD 578
DB 3598 ANDNVMAVVAQATDATV-----TIADDVLITDTSNIOGSTFSAVLTAGADATYLD 3651
QY 579 A 579
DB 3652 A 3652

RESULT 4

US-07-895-367-7

; Sequence 7, Application US/07895367

; GENERAL INFORMATION:

;; APPLICANT: Smit, John
;; APPLICANT: Bingle, Wade H
;; TITLE OF INVENTION: Bacterial surface protein expression
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Shlesinger, Arkwright & Garvey
;; STREET: 3000 South Eads Street
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: USA
;; ZIP: 22202

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/895,367
;; FILING DATE: 19920609
;; CLASSIFICATION: 424

;; ATTORNEY/AGENT INFORMATION:
;; NAME: Garvey, George A
;; REGISTRATION NUMBER: 17737
;; REFERENCE/DOCKET NUMBER: 5946
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-684-5600
;; TELEFAX: 703-836-5288

;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1026 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-07-895-367-7

Query Match 11.0%; Score 333.5; DB 3; Length 1026;
Best Local Similarity 24.7%; Pred. No. 1.1e-18;
Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

QY 16 GLKTAFTTSTAALMLTSGSVL-----GAATVTAD-CAELAAGT-----NIGP 58
DB 315 GIETMNTVSGAAILTNTSSGVTGLTALNTNTSGAOTVTAGAGONLTATTAQAANNVAV 374
QY 59 GAGAFV-----AGSTLQYTGAFVTDADVSRAALDNNFAAGLFSVTGDISL-----105
DB 375 DGGANVTASTGVTSGTIT--VCANSAASCTSVSVANSSTTTTGALVTTGTAIVTAQT 432
QY 106 -GSVD-----TGGANKLAVNIDGLTLTLTGTTAAYGANPALLFQCGQAAANN 154
DB 433 AGNAVNTLTQADVTVTGNSSTTAVTVTQ--TAAATAGATVAGRVNGAVTITDSAAASAT 490
QY 155 TTTALGNITLGGANAGLTIASDPDLGPITLAG-----NIDG 191
DB 491 TAGKIATVILGSPGAA-TI--DSSALTVNLSGTGSLGIGRGALPATPTANTLTUNVG 547
QY 192 ---GGIITDNTDAA-----INGTIGNTPAAQISIGASTLSGLGAVIKATTTKLTNA 240
DB 548 LITTGAITDSEAAADGFTTINAGTASSTIASLVAADATTILNISGDARVITTSATAA 607
QY 241 APVLTLTNA-----NAVLTGAVDNT-----260
DB 608 LTGITVNSVGATFLGAEATGLVFTGGAGADSILLGATTKAIYMGAGDDTVTVSSATLGA 667
QY 261 ---TGSD--DVGVLNGLALSQVTGNIGNTNSLATISVCAGTATLG-----G 302
DB 668 GGSVNGGDDTVLVANVNGSSFSADPAFGFETLR-----AGAAOGSHNANGFTALQLG 723
QY 303 AVIKATTTTKLTNA-----SVLTLTNAVLTGAVDNTTGGDNVGNLNGALSOVTGNIGN 357
DB 724 ATAGATT--FTNVAVNVGLTVLAAPTGTITVTLANATGTSDEVNLTLSAALAACTVA-780
QY 358 TNSLATINIGA-----GVATLDGAVIKATTTKLTDDASVLIFTNPVVVTVTAIDNTGNANKG 413

Db 781 LAGVETVNIATDNTTAHVDTLTLQATSAK-----SIVVTG-----NAG 820

Qy 414 VVIETGASTVDNIGTAVLAESVAGLLQIOGVVKANAINLTDNASVVFTGDSVT 473

Db 821 LNL-----TNTGNTA-----VTSFASAVTGTCSAVTFVSANTTV 855

Qy 474 GSIGTELFATVNIAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGN 532

Db 856 GEV-----VTIRGAGADSLTGSATANDTIIGGAGADTLVYTGCT-----DTFTGGT 902

Qy 533 NATL-NINAAGT-----VIANDVSIGTVAQINQNKIFVINAKNADVDILLDAQAISFKGA 587

Db 903 GADIFDINAIGTSTAFVITDAVNG-----DKLDLVGISTNGA 940

Qy 588 -ASRLFLANVSL 598

Db 941 IADGAFGAAVTL 952

RESULT 5

US-08-614-377-7

; Sequence 7, Application US/08614377

; GENERAL INFORMATION:

; APPLICANT: Smit, John

; APPLICANT: Bingle, Wade H.

; APPLICANT: Nomellini, John F.

; TITLE OF INVENTION: EXPRESSION AND SECRETION OF

; TITLE OF INVENTION: HETEROLOGOUS

; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson PC

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/614,377

; FILING DATE: 12-MAR-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/194,290

; FILING DATE: 09-FEB-1994

; CLASSIFICATION: 435

; APPLICATION NUMBER: US 07/895,367

; FILING DATE: 09-JUNE-1992

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Tsao, Y. Rocky

; REGISTRATION NUMBER: 34053

; REFERENCE/DOCKET NUMBER: 08106/002001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1026 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-614-377-7

Query Match 11.0%; Score 333.5; DB 10; Length 1026;

Best Local Similarity 24.7%; Pred. No. 1.1e-18;

Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

Qy 16 GLKTALEFTTAAIIMLTGSGVL-----GAARTVTAD-GAELAAAGT-----NITGP 58

Db 315 GIETMNTVSGAAITLNTSSGVTGLTALNTSGAAQVTVTAGAGQNLTTATTAQAANNVAV 374

Qy 59 GAGAFV-----AGSTLQVTGAFVTDADYSVRALDLNFAAGLFSVVTGDISL----- 105

Db 375 DGGANVTVASTGVTSCTTT--VGANSAASGTVSVSVANSSTTTTTCATIAVTGCTAVTVTAQT 432

Qy 106 -GSVVD-----TGGANKLAVNIDDDGLTLTLTGTTAAYGANPALLFGGGAANN 154

Db 433 AGNAVNTLTOADVTVTGNSSTTAVTVTQ--TAATAGATVAGRVNGAVTITDASAASAT 490

Qy 155 TYTALGNITLGGANAGLTIASDPDVLGPITLAG-----NIDG 191

Db 491 TAGKIATVTLGSPGAA-TI--DSSALTTVNLSGTSLGIGRGALTATPTANTLTLNVNG 547

Qy 192 ---CGIITDNTDAA-----INGTGNTPAAQISIGASTLSLGGAVIAKATTTKLTNA 240

Db 548 LTTTGAITDSEAAADGFTTINIAGSTASSTIASLVAADATTLNISGDARVTTISHTAAA 607

Qy 241 APVLTLTNA-----NAVLTGAVDNT----- 260

Db 508 LTGITVINSVGATLGAELATGLVFTGGAGADSILLGATTKAIVMGAGDDTVTVSSATLGA 667

Qy 261 ---TGGD--DVGVNLNGLSQTGNIGNTNLSLTIATSVGAGTATLG-----G 302

Db 668 GGSVNGGDTDLVANVNGSSFSADPAFGPFETLRV---AGAAQSGSHNANGFTALQLG 723

Qy 303 AVIKATTTKLTNAA-----SVLTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVVTGNIGN 357

Db 724 ATAGATT--FTNVAVNVGLTVLAAPTGTTVTLANATGTSDFNLTLSSSALAAGTVA- 780

Qy 358 TNSLATINIGA---GVATLDGAVIKATTTKLTDDASVLIFTPNPVVTGAIDNTGNANKG 413

Db 781 LAGVETVNIATDNTTAHVDTLTLQATSAK-----SIVVTG-----NAG 820

Qy 414 VVIETGASTVDNIGTAVLAESVAGLLQIOGVVKANAINLTDNASVVFTGDSVT 473

Db 821 LNL-----TNTGNTA-----VTSFASAVTGTCSAVTFVSANTTV 855

Qy 474 GSIGTELFATVNIAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGN 532

Db 856 GEV-----VTIRGAGADSLTGSATANDTIIGGAGADTLVYTGCT-----DTFTGGT 902

Qy 533 NATL-NINAAGT-----VIANDVSIGTVAQINQNKIFVINAKNADVDILLDAQAISFKGA 587

Db 903 GADIFDINAIGTSTAFVITDAVNG-----DKLDLVGISTNGA 940

Qy 588 -ASRLFLANVSL 598

Db 941 IADGAFGAAVTL 952

RESULT 6

US-08-194-290A-7

; Sequence 7, Application US/08194290A

; GENERAL INFORMATION:

; APPLICANT: Smit, John

; APPLICANT: Bingle, Wade H.

; TITLE OF INVENTION: BACTERIAL SURFACE PROTEIN EXPRESSION

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson PC

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSeq for Windows Version 2.0

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/194,290A
;; FILING DATE: 19-FEB-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/895,367
;; FILING DATE: 09-JUN-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Tsao, Y. Rocky
;; REGISTRATION NUMBER: 34,053
;; REFERENCE/DOCKET NUMBER: 08106/002001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-542-5070
;; TELEFAX: 617-542-8906
;; INFORMATION FOR SEQ ID NO. 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1026 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FRAGMENT TYPE: internal
;; US-08-194-290A-7

Query Match 11.0%; Score 331; DB 5; Length 1026;
Best Local Similarity 27.3%; Pred. No. 1.8e-18;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

```
QY 23 TTSTAAILMTGSGVLGAART-----VTADG-----AELAAAGTNIGPGA 60
Db 413 TTTTGAIVTGGTAVTVAQTAGNAVNNTLTQADVTVTGNSSTAVTQTAAAT-----A 467

QY 61 GAFVAGSTLQYTGAFVTDADVS-----VRALDINFAAGL-----FSVTG-- 101
Db 468 GATVAG--RVNGAVITDSSAASATTACKIATVILGSFGAATIDSSALTIVNLSGTGFS 524

QY 102 -DISGVSVDGTGANKLAVNIDDLTLTGCTGTAAYGANPALLFGGQAAANNYYTALG 160
Db 525 LGIGRGALTATPTANTLTINVNG---LTTTGALT-----DSEAAADDDGFT--- 566

QY 161 NITLGGANAGLTIAS----DPDVLGPITLAGNIDGG--IITDNTDAAING-TIGNTPA 213
Db 567 TINIAGSTASSIASLVAADATTL-----NISGDARVTIISHTAAALTGITVTVNSVGA 619

QY 214 ---AQISIGASTLSLGG-----VIKATTKLTNAAPVLTNANAVLTGAVDNTTGGD- 264
Db 620 TLGAELATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSATL-GAGGSVNGDG 676

QY 265 -DVGVLNGLALSOVVTGNIGNTNSLATISVGAGTATLG-----GAVIKATTK 311
Db 677 TDVLVANVNGSFSADPAFGFETLRV-----AGAAAGSHNANGFTALQIGATAGATT-- 730

QY 312 LTNA-----SVLTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVVTGNIGNTNSLATINI 366
Db 731 FTVNAVNVGLTVLAAPTGTITVTLANATGSDVFNLTLSAALAAGTVA-LAGVETVNI 789

QY 367 GA----GVATLDGAVIKATTKLTDDASVLIFTNPVVVTGAIDNTGNANKGVVIFTGAST 422
Db 824 --TNTGNTA-----VTSFDSAVTGTAPAVTFVSANTTVGEV----- 858

QY 483 ATVNICAGITLRAGGSIAANNIDFGA-ASNLEPENGAGKNYNLIGTIANGNNATL-NINA 540
Db 859 VTIIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DFTTGTGADIADIFDINA 911

QY 541 AGT-----VIANDVSGITVAQINQNNKIFVINAKNADVDILDAQAISFKGA-ASRFLAN 595
VSL 598
|:|
```

Db 950 VTL 952

RESULT 7

US-09-379-931-7
; Sequence 7, Application US/09379931
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: Nomellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CA
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 11.0%; Score 331; DB 17; Length 1026;
Best Local Similarity 27.3%; Pred. No. 1.8e-18;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

```
QY 23 TTSTAAILMTGSGVLGAART-----VTADG-----AELAAAGTNIGPGA 60
Db 413 TTTTGAIVTGGTAVTVAQTAGNAVNNTLTQADVTVTGNSSTAVTQTAAAT-----A 467

QY 61 GAFVAGSTLQYTGAFVTDADVS-----VRALDINFAAGL-----FSVTG-- 101
Db 468 GATVAG--RVNGAVITDSSAASATTACKIATVILGSFGAATIDSSALTIVNLSGTGFS 524

QY 102 -DISGVSVDGTGANKLAVNIDDLTLTGCTGTAAYGANPALLFGGQAAANNYYTALG 160
Db 525 LGIGRGALTATPTANTLTINVNG---LTTTGALT-----DSEAAADDDGFT--- 566

QY 161 NITLGGANAGLTIAS----DPDVLGPITLAGNIDGG--IITDNTDAAING-TIGNTPA 213
Db 567 TINIAGSTASSIASLVAADATTL-----NISGDARVTIISHTAAALTGITVTVNSVGA 619

QY 214 ---AQISIGASTLSLGG-----VIKATTKLTNAAPVLTNANAVLTGAVDNTTGGD- 264
Db 620 TLGAELATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSATL-GAGGSVNGDG 676

QY 265 -DVGVLNGLALSOVVTGNIGNTNSLATISVGAGTATLG-----GAVIKATTK 311
Db 677 TDVLVANVNGSFSADPAFGFETLRV-----AGAAAGSHNANGFTALQIGATAGATT-- 730

QY 312 LTNA-----SVLTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVVTGNIGNTNSLATINI 366
Db 731 FTVNAVNVGLTVLAAPTGTITVTLANATGSDVFNLTLSAALAAGTVA-LAGVETVNI 789

QY 367 GA----GVATLDGAVIKATTKLTDDASVLIFTNPVVVTGAIDNTGNANKGVVIFTGAST 422
Db 790 AATDTNTTAHVDTLTQATSAK-----SIVVTG-----NAGLNL----- 823

QY 423 VTDNIGNTAVLAESVVGAGLLQIQGGVVKANALINDNASVVTFTGDSVTGSGIGTELF 482
Db 824 --TNTGNTA-----VTSFDSAVTGTAPAVTFVSANTTVGEV----- 858

QY 483 ATVNICAGITLRAGGSIAANNIDFGA-ASNLEPENGAGKNYNLIGTIANGNNATL-NINA 540
Db 859 VTIIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DFTTGTGADIADIFDINA 911

QY 541 AGT-----VIANDVSGITVAQINQNNKIFVINAKNADVDILDAQAISFKGA-ASRFLAN 595
```

Db 912 IGTSTAFVTITDAAVG-----DKLDLVGISTNGAIDAGFAGAA 949
QY 596 VSL 598
Db 950 VTL 952

RESULT 8
US-09-743-731-5
; Sequence 5, Application US/09743731
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; TITLE OF INVENTION: CLEAVAGE OF CAULOBACTER PRODUCED
; FILE REFERENCE: 08106-004001
; CURRENT APPLICATION NUMBER: US/09/743,731
; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: PCT/CA99/00637
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: CA 2,237,704
; PRIOR FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-743-731-5

Query Match 11.08; Score 331; DB 21; Length 1026;
Best Local Similarity 27.38; Pred. No. 1.8e-18;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

QY 23 TTSTAAMLTGSGVLGAART-----VTADG-----AELAACTNIGPGA 60
Db 413 TTTTGALAVTGGTAVTVAQTAGNAVNNTLTQADVTVTGNSETTAVTVTQAAAT-----A 467

QY 61 GAFVAGSTLQYTGFTVDADVS-----VRALDLNFAAGL-----FSVTG-- 101
Db 468 GATVAG---RVNGAVTITDASAASATTAGKATVATVTLGSFGAATIDSSALATVNLSGTGS 524

QY 102 -DISLGSVVDTPGGANKLAVNIDDLTLTGTGTAAAYCANPALLFQGGQAAANNITYALG 160
Db 525 LGIGRGALTATPTANTLTVNG---LTTGATV-----DSEAAADGFT--- 566

QY 161 NITLGGANAGLTIAS---DPVLGPTTLACNIDGG--IITDNTDAING--TIGNTPA 213
Db 567 TINAGSTASSTIASLVAADATTL-----NISGDARVITTSHTAAALGITVTVNSVGA 619

QY 214 --AQISIGASTLSLGA---VIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264
Db 620 TLGHELATG--LVFTGGAGRDSILLGATTKAIVNGAGDDTVTVSSATL-LGAGGSVNGGDG 676

QY 265 -DVGVLNGLALSQVTGNIGNTNSLATISVCAGTATLG-----GAVIKATTK 311
Db 677 TDVLVANVNGSSFSADPAFGFELRV---AGAAAGSHNANGFTALQICATAGAT-- 730

QY 312 LTNA---SVLPLTNAVLTGAVDNTTGGDNGVNLGALSQVTGNIGNTNSLATINI 366
Db 731 FTNAVAVNGLTVLAAPPTGTTTTLANATGTSDFENLTSSSAAALAACTVA-LAGVETVNI 789

QY 367 GA---GVATLDGAVIKATTKLTDDASVLFTNPVVVTTCAIDNTGNANKGVVIFTGAST 422
Db 790 AATDNTTAVHDTLTQATSAK-----SIVVTG-----NAGLNL----- 823

QY 423 VTDNIGNTAVLAESVSGAGLQIQGGVYKANALNTDNASVTTFGDSTVTGSGTGTELF 482
Db 824 --TNTGNTA-----VTSFSDASAVTCAFAVTFVSANTTVGEV----- 858

QY 483 ATVNIGAGITLRAGGSLAANNIDFGA--ASNLEFNGPAGKNYNIIGTIANGNNATL-NINA 540
Db 859 VTRGGAGADSLGTSATANDTIIGGAGADTLVVTGGT-----DPTGTGCTGADIFDINA 911

QY 541 AGT-----VIANDVSGTVAQINQNNKIFVINAKNADVDILDAQAISPKGA-ASRLFLAN 595
Db 912 IGTSTAFVTITDAAVG-----DKLDLVGISTNGAIDAGFAGAA 949

QY 596 VSL 598
Db 950 VTL 952

RESULT 9
US-09-739-449-13178
; Sequence 5, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 13178
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-13178

Query Match 10.78; Score 323.5; DB 21; Length 1242;
Best Local Similarity 25.08; Pred. No. 1e-17;
Matches 163; Conservative 90; Mismatches 195; Indels 205; Gaps 36;

QY 20 ALFTTSTAAMLTGSGVLG---AARTVTADGAELAAG-TNIG-----PG----- 59
Db 233 AKLVVEAAAVTSGEALIGHRSASEATVGDGSKWTTGDLOVGGDTSDPGLAGNGLNV 292

QY 60 -AGAFVAGSTLQYTGAF-----TVTDADVSVALDLNLF-----AAGLFSVTDGDISLGS 107
Db 293 TAGGSV-DSTVAHLGVVAGATGSAIVDGKGSVMTVDRNSLEVSGVSGAGSLAVTG---GG 347

QY 108 VVD-----TCGANKLAVNIDDDGLTLTGTGTAAAYCANPALLFQGGQAAANNITYA 158
Db 348 LVDAANIIIGTNTGNGSVRVSGADSTVSKRSLNVLGNGSVMTVBAAGAVKSRDGYA 407

QY 159 LGNITLGGANAGLTIASDPDVLGPITLAGNIDGGIITDNTDAAINCT--IGNTNPAQI 216
Db 408 ---TYGGSISAVTVTG-----DGSWAMTGTFFGVYASGAT-- 440

QY 217 SIGASTLSLGGAV--IKATTTKLTNAAPVLTLTNANAVLTGAVDNTT---GGDDVGVNL 271
Db 441 --GNVTVSNGGATRAATGVTGLDLAGASGTTITGAGSKVTAYVDNGTVNSGSDVGVFGQ- 497

QY 272 NGALSQVTGN-----IGNTNSLATISVG-----AGTATL---GGA 303
Db 498 SGLSVVNGSGLDAYNLYVGNALGSSGAVLVSGVSHVSDGLMVVGNAGNGSVETGGA 557

QY 304 VIKATTTKL-TNAASVLTLT-----NAVLTGAVDNTT---GGDNVGVY----- 342
Db 558 SLAAPTILATEAGSTCVLSIGAGSGGTARSAGAVEARAIAFGANGSGSVFNHSETGYTL 617

QY 343 --NLGALSQVTGNIGNTNSLATINIGAGVATLDGAVIKATTK-----LTDASVLI-- 393
Db 618 SADISGA-GRVAEAGVT-TLGNNSYSGGTTISAGMLKGTAKSFSGSGGVVNNAEVLVDG 675

QY 394 ---FTNPVVVTGAIDNTGNANKGVVLTFTGASTVTDNIGTAVLAESVSGAGLQIQGGV 450
Db 676 GGTLSNAISGTSGFEKTGDN---LLLTGNSTYS---GATA-----VSAGKLSVNGSL- 722

QY 451 KANALNTDNASVVTFTGDSV---TGSIGGTETELFATVNIGAGITLRAGGSLAANNIDFGA 508
Db 723 -----ASAVSVGSGATVGGTGIG-----GLTVNSGGTFLAPGN---S 756

QY 509 ASNLEFNGPAGKNYLNIGTIANGNNATLINAAGT-----VIANDVSI 551
 Db 757 ICTLTSTGNA-----TFAGSGTYAVEIDADGSSDRLAVTGTITIANDVSL 801

RESULT 10
 US-09-803-110-13178
 ; Sequence 13178, Application US/09803110
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15490)D
 ; CURRENT APPLICATION NUMBER: US/09/803,110
 ; CURRENT FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/739,449
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 09/514,000
 ; PRIOR FILING DATE: 2000-02-23
 ; PRIOR APPLICATION NUMBER: US 60/168,139
 ; PRIOR FILING DATE: 1999-12-01
 ; NUMBER OF SEQ ID NOS: 13351
 ; SEQ ID NO 13178
 ; LENGTH: 1242
 ; TYPE: PRT
 ; ORGANISM: Agrobacterium tumefaciens
 ; US-09-803-110-13178

Query Match 10.7%; Score 323.5; DB 22; Length 1242;
 Best Local Similarity 25.0%; Pred. No. 1e-17;
 Matches 163; Conservative 90; Mismatches 195; Indels 205; Gaps 36;
 QY 20 ALFTTSTAALMLTGSVVG-----AARTVTADGAELAAAG-TNIG-----PG----- 59
 Db 233 AKLVVEAATVTSGEAIIHRHSAEATVTGDSKWTGTGDLQVGGTSDPGGLAGNLTNV 292
 QY 60 -AGAFVAGSTLQYTGAF-----TVTDADVSVRALDNNF-----AAGLFVSGTGLSLS 107
 Db 293 TAGGSV-DSTVAHLGVVAGATGSAIVDGGKSVWTVDRNSLEVGVSGAGSLAVTG-----GG 347
 QY 108 VVD-----TGGANKLAVNIDDLTLTGTGTAAAGANPALLFQGGQAAANTTYA 158
 Db 348 LVDAANIIITGNTGNGSVRVSGADSTVKRSRLNVLGXGSMVTVEAGGAVKSRDGYVA 407
 QY 159 LGNITLGLGAGNAGLTASDPDLVLPITLAGNIDGGIITDNTDAIINGT--IGNTPNPAOI 216
 Db 408 -----TYGGSTSAVTG-----DGSWAMTGTFFGVASGAT-- 440
 QY 217 SIGASTLSLGGAV--IKATTTKLITNAAPVLTITNANAVLTGAVDNTT---GGDDVGVNML 271
 Db 441 --GNVTVSNGGAIKATGVTGLDLAGAGTWTITGAGSKVTAYVDNGTVNSGSDVGVFGQ- 497
 QY 272 NGALSOVTGN-----IGNTNSLATISVG-----AGTATL---GGA 303
 Db 498 SGSLSVYNGGSLDAYNLYVGNALGSSGAVLVSGVSHVSDGLVMVYGNAGNGSVETGGA 557
 QY 304 VIKATTKL-TNAASVLTIT-----NAVLTGAVDNTT---GGDNVGVV----- 342
 Db 558 SLAAPILLIATEAGTGVLSIGAGSCQARSAGAVEARAIFAGAGNGSIVFNHSETGYTL 617
 QY 343 --NLSGALSQVTGNIGNTNSLATINIGAGVATLDGAVIKATTK-----LTDASVLI-- 393
 Db 618 SADISCA-GRVVAEAGVT-TLSCNNYSVGGTTISAGMLKGTAKSFSGSGGVVNAELVVDG 675
 QY 394 ---FTNPVVVTGADITNGANKGVVLTGASIVTDNIGNTAUVAEVSAGLLOIOGGVV 450
 Db 676 GGTLSNAISGTSFERTGDCN---LLLTGNSTYS---GATA-----VSAGKLSVNGSL- 722
 QY 451 KANAILNDAASVVTFTGDSV--TGSIGGTGELFATFVNIGAGITTLRAGGSLAANNIDFGA 508
 Db 723 -----ASAVSVGSGATVGGTGIG-----GLTVNSGGTLAPGN-----S 756
 QY 509 ASNLEFNGPAGKNYLNIGTIANGNNATLINAAGT-----VIANDVSI 551

Db 757 ICTLTSTGNA-----TFAGSGTYAVEIDADGSSDRLAVTGTITIANDVSL 801
 RESULT 11
 US-09-167-568-61
 ; Sequence 61, Application US/09167568
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: KLEIN, Michael H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
 ; TITLE OF INVENTION: INFLUENZA HIGH MOLECULAR WEIGHT PROTEINS
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/167,568
 ; FILING DATE: 07-OCT-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 61:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1180 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-167-568-61

Query Match 10.7%; Score 322; DB 15; Length 1180;
 Best Local Similarity 25.2%; Pred. No. 1.3e-17;
 Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;
 QY 48 AELAACTNIGPGAGAF-VAGSTLQYTGAFV-----TDADVSVRALDNNFAAGLFSVT 100
 Db 543 AEIQIGNISQREGNLTISDDKINIKQITIKKGVNGENSDDSTKS-----QANLTIKT 596
 QY 101 GDISLGSVVDGANKLAVNIDDLTLTGTGTAAAGANPALLF-----OGGQA 150
 Db 597 KELKLTQDLNISGNKAKIVAKDSSNLTIGNSDDSGNTSAKTVTFNNVKDSKISADGHV 656
 QY 151 AANNVYALG---NITLGA--NAGLTI-ASDPDLVLPITLAGNID---GGGIITDNTDA 201
 Db 657 TLNSKYKTLSDNDNTEGGSDNNTGLTITAKDVEVNNITSHKTVNVSAANGGITTKGT 716
 QY 202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLITNAAPVLTITNANAVLTGAVDNTT 261
 Db 717 TINATAGN-----VEITAHTGSIQGI-----ESKP-----GSVTIVA 749
 QY 262 GGDVGVNLSGALSQVTGNIGNTNSLA-----TISVGAGTATLGGAV-----IKA 307
 Db 750 GGDTLAVGNISGNNAVTVTANSALTTLAGSTIKGTESITTSQSGNIGKISGKTVNVKA 809
 QY 308 TTTKLNAASVLTITLNAV-----TCGAVDNTTGGDNVGVNLSGALSQVTGNIGNTNSLA 362
 Db 810 TNSLTQAQDSKIBATEGEANVTSKTSIIIGTISGGTVEVTATGLTQAQSTITGTESVT 869

```

363 QY 363 TI-----NIGAVATLCAVIAKATTK--LTDDASVLITF-----NPVVVTGAIDNTGNANK 412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
870 Db 870 TSSQSGNIG---GMISGGKVEYSATKDLITTKSGSEIKATAGEVNVTSATGTIDGTISGNT 926
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 QY 413 -GVVIETGASTVINDIGNTAVLAESVSGAGLLOIOGG--VVKANAINLTDNASVVTFITGD 469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
892 Db 927 VNVNTANTGDLTVED-----AAKIDATGGAATLITATSGKLTTKASSITTSANNQVNSAKD 981
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 QY 470 STVTSGIGGTETFPATVNIAGAGITLRAGGSLAANNIDFGA----ASNLEFNGPAGKNYNLI 525
      :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
982 Db 982 GSIGGGINAANV--TLNTTTCALLTTVKGSSINANS---GTLVINAKDAELNGEASGNHTVV 1036
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
526 QY 526 GTIANGNNATLNTNAGTIVANDSVISGTVAQINIQNNKIFIVINAKNADVDLIDQAASIFK 585
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
810 Db 1037 -----NAT-NANGSGSVIAT-----TSSRVITGTD-LITINGLN-----IISK 1072
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 QY 586 GAASRLFLANVSLQMLELSLLKIIPVLLTV 616
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
810 Db 1073 NGINTVLLKGVKID-----VKYIQGIASV 1097
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-167-568A-61
; Sequence 61, Application US/09167568A
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: KLEIN, Michael H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
; TITLE OF INVENTION: INFLUENZAЕ HIGH MOLECULAR WEIGHT PROTEINS
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1K7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/167,568A
; FILING DATE: 07-OCT-1998
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-167-568A-61

Query Match 10.7%; Score 322; DB 15; Length 1180;
Best Local Similarity 25.2%; Pred. No. 1.3e-17;
Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps

QY 48 ABLAAGTNTGPGAGAF-VAGSTLAYTGAFV-----TDADVSVRALDLNNFAAGLFSVT 100
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 AEIQIGNITSQEGNLTISSDKINIKQITIKKGVNGENSDSSTKS-----QANLTIKT 596
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 GDISLGSVVDGKANKLAVNIDDLGLTLTLTGTAAYCANPALFF-----OGGQA 150
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 KELKTQDLNTSGFNKAKIVAKDSNLTIGNSDDSGNTSAKTVTFNNVKDKSIADGKVV 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY	151	AAANTYTAGL---NITLGG--NAGLTI--ASDPDVLGPITILAGNID---GGGIITIDND	201
DB	657	TLASKVKTLSDNDNTEGGSDNNTGLITAKDVEYNNNTISHTKTVNVSAAANGGITKTGT	716
QY	202	AINGTIGTNPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTLTANAVLTGAVDNTT	261
DB	717	TINATAGN---VEITAHTSIOGGI-----ESKP-----GSVTIVA	749
QY	262	GGDDVGVNLNGLALSQVOTGNIGNTSLA-----TISVGAGTATLGGAV-----IKA	307
DB	750	GGDTLAVGNISGNVAVTVTANSALITLAGSTIKGTESITTSQSGNIGGKISGKTVNVKA	809
QY	308	TTTKLNTNAAVLTLTNAVLI-----TCADVNTGGDNVGVNLSGALSQVOTGNIGNTSLA	362
DB	810	TNSLTTLQAQSKIBATEGEANVTSKTSIIIGTISGGTVEVTATEGLTTQAGSTITGTESVT	869
QY	363	TI-----NIGAGVATLGGAVIKATTK--LTDDASVLIFT---NPVVVTGAIDNTGNANK	412
DB	870	TSSQSGNIG---GMISGGKVEVSATKDLITKSGSIIKATAGEVNTVSATGIDGTISGNT	926
QY	413	-GVVIFTGASTVTDNIGNTAVLAEYSVSGAGLQIOGG--VVKANAINLTDNASVVVTETGD	469
DB	927	VNVTANTGDLTVED-----AAKIDATGGAATLTATSGKLTTKASSITSANNQVNLSAKD	981
QY	470	STVTSIGGTGELFATVNICAGITLRAGSILAANNIDFGA----ASNLEFNGPAGKNTNLI	525
DB	982	GSITGGGINAANV--TLNPTGALTTVKGSINANS---GTLVINAKDAELNGEASGNHTV	1036
QY	526	GTIANGNNATLINAAGTVIANDSVSTGTVQAINQNNKIFVINAKNADVDILDAQAISFK	585
DB	1037	-----NAT-NANGSGSVIAT-----TSSRVNITGD--LIINGLN-----IISK	1072
QY	586	GAASRLFLNWSLQMLELSLLKIIPVLLTV	616
DB	1073	NGINTVLLKGVKID-----VKYIQPGIASV	1097

RESULT 13
 US-09-167-568-59
 ; Sequence 59, Application US/09167568
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
 ; TITLE OF INVENTION: INFLUENZAЕ HIGH MOLECULAR WEIGHT PROTEINS
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/167,568
 ; FILING DATE: 07-OCT-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-828 MJS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 59:
 ; SEQUENCE CHARACTERISTICS:

RESULT 13
 US-09-167-568-59
 ; Sequence 59, Application US/09167568
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
 ; TITLE OF INVENTION: INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/167,568
 ; FILING DATE: 07-OCT-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-828 MCS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 59:
 ; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1188 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-167-568-59

Query Match 10.7%; Score 322; DB 15; Length 1188;
Best Local Similarity 25.2%; Pred. No. 1.3e-17;
Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;

QY 48 AELAACTNIGCAGAF-VAGSTLQYTGATV-----TDADVSVRALDNNFAAGLSFVT 100
DB 551 AEIQGGNISQEGNLTSSDKINITTKQIKKGVNGENSSTKS-----QANLTIKT 604
QY 101 GDISGVVDYTGANKLAVNDDGLTLTGTAAYGANPALLF-----QGGQA 150
DB 605 KELKUTQDLNISGPNKAKIVAKDSSNLTIGNSDDSGNTSAKTVTFNNVKDSKISADGHV 664
QY 151 AANNTYTALG---NITLGA--NAGLTI--ASDPDVLGPITLAGNID---GGIITDNTDA 201
DB 665 TLNSKVKTLSDNDNTEGGSDNNTGLITAKDVEYNNITSHKTVNVSAAANGIITKTGT 724
QY 202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTNTNANAVLTGAVDNTT 261
DB 725 TINATAGN-----VEITAHTSGIOGGI-----ESKP-----GSVTIVA 757
QY 262 GDDVGVNLNLCALSOVTCNIGNTNSLA-----TISVGAGTATLGGAV-----IKA 307
DB 758 GGDTLAVGNISGNAVTVTANSALTTLAGSTIKGTESITTSQSGNIGKISGKTVNVKA 817
QY 308 TTTKLTNAASVLTLTNAVL-----TCAVDNTTGGDNGVNVNLSGALSQVGTGIGNTNSLA 362
DB 818 TNSLTQADSKLEATEGEANVTSKTSIIGGTISGGTVEVTATEGLTQAGSTITGESVT 877
QY 363 TI-----NIGAGVATLDGAVIKATTK--LTDDASVLIFT---NPVVVTGAIDNTGNANK 412
DB 878 TSSQSGNIG---GMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTDIGTISGNT 934
QY 413 -GVVIFTGASTVTDNIGNTAVLAESVAGLLQIOGG--VVKANALNTDNASVVTFTGD 469
DB 935 VNVNTANGDLTVED-----AAKIDATGGAATLTATSGKLTTRKASSITTSANNQVNLK 989
QY 470 STVTSGIGTELFATVNIAGITLRAGGSLAANNIDFGA---ASNLEFNGPAGKYNLI 525
DB 990 GSIGGNINAANY--TLNITGALTTVKGSINANS---GFLVINAKDAELNGEASGNHTVV 1044
QY 526 GTIANGNATLINAGTVIANDVSGTVAQINIQNNKIFVINAKNADVDILDAQAISPK 585
DB 1045 -----NAT--NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----ITSK 1080
QY 586 GAASRLFLANVSLQMIELSLKIIYPVLLTV 616
DB 1081 NGINTVLLKGVKID-----VKYIQPGIASV 1105

RESULT 14
US-09-167-568A-59
; Sequence 59, Application US/09167568A
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
; TITLE OF INVENTION: INFLUENZA HIGH MOLECULAR WEIGHT PROTEINS
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/167,568A
; FILING DATE: 07-OCT-1998
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-828 MIS:j:b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-167-568A-59

Query Match 10.7%; Score 322; DB 15; Length 1188;
Best Local Similarity 25.2%; Pred. No. 1.3e-17;
Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;

QY 48 AELAACTNIGCAGAF-VAGSTLQYTGATV-----TDADVSVRALDNNFAAGLSFVT 100
DB 551 AEIQGGNISQEGNLTSSDKINITTKQIKKGVNGENSSTKS-----QANLTIKT 604
QY 101 GDISGVVDYTGANKLAVNDDGLTLTGTAAYGANPALLF-----QGGQA 150
DB 605 KELKUTQDLNISGPNKAKIVAKDSSNLTIGNSDDSGNTSAKTVTFNNVKDSKISADGHV 664
QY 151 AANNTYTALG---NITLGA--NAGLTI--ASDPDVLGPITLAGNID---GGIITDNTDA 201
DB 665 TLNSKVKTLSDNDNTEGGSDNNTGLITAKDVEYNNITSHKTVNVSAAANGIITKTGT 724
QY 202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTNTNANAVLTGAVDNTT 261
DB 725 TINATAGN-----VEITAHTSGIOGGI-----ESKP-----GSVTIVA 757
QY 262 GDDVGVNLNLCALSOVTCNIGNTNSLA-----TISVGAGTATLGGAV-----IKA 307
DB 758 GGDTLAVGNISGNAVTVTANSALTTLAGSTIKGTESITTSQSGNIGKISGKTVNVKA 817
QY 308 TTTKLTNAASVLTLTNAVL-----TCAVDNTTGGDNGVNVNLSGALSQVGTGIGNTNSLA 362
DB 818 TNSLTQADSKLEATEGEANVTSKTSIIGGTISGGTVEVTATEGLTQAGSTITGESVT 877
QY 363 TI-----NIGAGVATLDGAVIKATTK--LTDDASVLIFT---NPVVVTGAIDNTGNANK 412
DB 878 TSSQSGNIG---GMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTDIGTISGNT 934
QY 413 -GVVIFTGASTVTDNIGNTAVLAESVAGLLQIOGG--VVKANALNTDNASVVTFTGD 469
DB 935 VNVNTANGDLTVED-----AAKIDATGGAATLTATSGKLTTRKASSITTSANNQVNLK 989
QY 470 STVTSGIGTELFATVNIAGITLRAGGSLAANNIDFGA---ASNLEFNGPAGKYNLI 525
DB 990 GSIGGNINAANY--TLNITGALTTVKGSINANS---GFLVINAKDAELNGEASGNHTVV 1044
QY 526 GTIANGNATLINAGTVIANDVSGTVAQINIQNNKIFVINAKNADVDILDAQAISPK 585
DB 1045 -----NAT--NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----IISK 1080
QY 586 GAASRLFLANVSLQMIELSLKIIYPVLLTV 616
DB 1081 NGINTVLLKGVKID-----VKYIQPGIASV 1105

RESULT 15
```

US-09-167-568-35
 ; Sequence 35, Application US/09167568
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
 ; TITLE OF INVENTION: INFLUENZA HIGH MOLECULAR WEIGHT PROTEINS
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto,
 ; STATE: Ontario,
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/167,568
 ; FILING DATE: 07-OCT-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 35:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 915 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-167-568-35

Query Match 10.5%; Score 316.5; DB 15; Length 915;
 Best Local Similarity 23.8%; Pred. No. 2.7e-17;
 Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

QY 42 TVTADGAELAGTNGPCAGAF-VAGSTLQYTGAFVTDADVSVRALDLNN-FAAGLFV 99
 DB 229 TNGDNTFQIGGNISQKEGNLTSSDKVNITERITI-KAGVNGDSDSNEATSANLTIK 287
 QY 100 TGDISLGSVVDGTGGANKLAVNIDDLTLTGTCTAAYGANPALLFOGGQAAANNYYTAL 159
 DB 288 TKELKLTNDLNSGFNKAETAKDNSNLT-----GDNSDAGNTDAK- 329
 QY 160 GNITLGGANAGLTASDPVLGPITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIG 219
 DB 330 -KVTFSNVKDSKISASDHNV---TLNKKVETSG-DTSDTEGCGNNNTGLTITAKNVTN 383
 QY 220 ASTLSLGGAVIKATTKLTNAAPVLTITNANVLTVGAVDNTGDDVGVNLNGLALSQVT 279
 DB 384 NNITSHKVTNITASENVTTKAGTTINATGTSVEVTAKTGKIGGIESNGSNVNITASGDT 443
 QY 280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTKLTNAASVLTITNAV-----LTCAVD 331
 DB 444 LNVSNITQNVTVAAASGAVTTTKGTINAT---TGNANITTKTGEINKEVKSASGNV 499
 QY 332 NTGGDNVGVNLSGALSQVTVGNIGNTNSLATINIGAGVATLDGAVIKATTKLTDDASV 391
 DB 500 ITASGNTLVNSNITQNVTVTANS-----AITTEGSTINAT---TGDANI 543
 QY 392 LIFTNPVVVTGADNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVCGALLOIQ--- 446
 DB 544 -----TTQGNINGKVESGSSGVTIATQTLAVNIGSDTV--TITADKGLTTQTSS 595
 QY 447 -----GGVVKANAINLTDNASVVTFTGD-----STVTGSIGG 478

Db 596 KINGTKSVTTSSQSGDISGTISGNTVSVSATGSLTTQAGSKTEAKTGEANVTSA GTIGG 655
 QY 479 TELFATVNI-----GAGITLRA-GGSLAANNIDFGAASNLEFNGPAG 519
 Db 656 TISGNTVNVNTANDNLTIKDGARIKATGAVTLTATGTLTETSSDITSSNGQTTLTA- 714
 QY 520 KYNLIGTIANGNNATLNINAAAGTVIAN---DVSIGTVAQINIQNNKIFVINAKNADVD- 575
 Db 715 KDSSIAGSI-NAANVTLANTTGTLTVAGSKIEAASGTL-----VINAKQAQLDG 762
 QY 576 -----ILDAQAISFKGA--ASRLFLANVSLQMIELSLKII 609
 Db 763 AASGDHTVVNATNANGSGSVIATTSRNVNITGDLITINGLNII 805

Search completed: April 14, 2003, 16:14:42
 Job time : 165 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:09:29 ; Search time 82 Seconds
(without alignments)
1176.179 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISKLKFORAIQKGLKTA.....MIELSLKLIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 704720 seqs, 155810517 residues

Total number of hits satisfying chosen parameters: 704720

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New.*

1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	392	13.0	1862	6	US-10-282-122A-49757
2	373	12.3	979	6	US-10-282-122A-47372
3	366.5	12.1	1331	6	US-10-282-122A-47930
4	362	12.0	2457	6	US-10-282-122A-49854
5	353.5	11.7	827	6	US-10-282-122A-47312
6	350	11.6	4327	6	US-10-369-493-10178
7	342.5	11.3	3286	6	US-10-282-122A-49697
8	341.5	11.3	1407	6	US-10-282-122A-47709
9	331	11.0	1026	6	US-10-223-597-7
10	329.5	10.9	3705	6	US-10-282-122A-77944
11	322	10.7	1180	6	US-10-193-764-61
12	322	10.6	3073	6	US-10-193-764-59
13	319.5	10.5	915	6	US-10-193-764-35
14	316.5	10.5	1222	6	US-10-193-764-37
15	316.5	10.5	1228	6	US-10-193-764-34
16	316.5	10.4	1129	6	US-10-282-122A-48048
17	315.5	10.2	2056	6	US-10-282-122A-50912
18	308	10.1	1074	6	US-10-282-122A-50616
19	306.5	10.0	1215	6	US-10-369-493-10073
20	301	9.9	3013	6	US-10-282-122A-78257
21	298	9.7	969	6	US-10-193-764-32
22	293.5	9.7	975	6	US-10-193-764-30
23	293.5	9.7	1465	6	US-10-369-493-13955
24	293	9.7	6310	6	US-10-282-122A-67793
25	293	9.6	1220	6	US-10-193-764-28
26	289	9.6	1220	6	US-10-193-764-28

Sequence 26, Appl
Sequence 47521, A
Sequence 69, Appl
Sequence 67, Appl
Sequence 50634, A
Sequence 70017, A
Sequence 48220, A
Sequence 31, Appl
Sequence 54678, A
Sequence 65, Appl
Sequence 63, Appl
Sequence 28, Appl
Sequence 49773, A
Sequence 41, Appl
Sequence 39, Appl
Sequence 78305, A
Sequence 18460, A
Sequence 55546, A
Sequence 57, Appl

ALIGNMENTS

RESULT 1

US-10-282-122A-49757
; Sequence 49757, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49757
; LENGTH: 1862
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49757

```
Query Match      13.0%; Score 392; DB 6; Length 1862;
Best Local Similarity 28.0%; Pred. No. 5.5e-15;
Matches 161; Conservative 66; Mismatches 197; Indels 152; Gaps 26;

Qy 13 IOKGLKTAFTTSTAAILMTGSGVGLGAARTVTADGAELAAAGTNIQPGAGAFVAGSTLQYT 72
Db 1348 ITTGDLDSL-GASTSAV---GDVTLNAGRNA-----LNGTLVGGNGYISAGQDVTPG 1397

Qy 73 GAFTVTDADVSRALDLNFAAGLFSVTGDISLGSVDVTGGANKLAVNIDDLGLTLTLTGT 132
Db 1398 GTQAFTHAVAL-----GAGGDIALTGLSLQANSVQATGGDNRALNNVTSTTLTLTAN 1449

Qy 133 GTAAYGANPALLFOGQQAANNVTYALNITLGGANAGLTIASDPDVLGPTITLAGNIDGG 192
Db 1450 GNAG-----NGDAITGTATAPGLVTLNAR-----DVL-----ISGALGGG 1486

Qy 193 GIIT--DNTDAAINCT---IGNTPAAOISIGASTLSLGGAVIKATTTKLTNAAAPVLT 247
Db 1487 TTVTLAQRNATVSGTVOAGVDL-----VTAT 1514

Qy 248 NANAVLTGAVDNTTGGDDVGVNLNGAL--SQVTGNIGNTNSLTIASVG----- 294
Db 1515 NGSAAITGTATTT-----GALNVTSGLDLTQGGQIAVTNSL-TVQACTDVELTGTVAQ 1567

Qy 295 -AGTATLGAIVKATTTKLTNAAVSLTTLTNAVLTGAVD-----NTTGGDNVGVNLSGALS 349
Db 1568 AAGSLTAGQDITGGGSAFAQAATLSAANNVALTGALQSGISVTGGNAGL-----GSAQ 1623

Qy 350 QVTGNIGNTNSLATINICAGVATLDGAV--IKATTTKLTDDASVLITFPVVTGAIDNT 407
Db 1624 ATTGNLDVT---ANGNAGQDVTTGGVATAGTTTLQARVDS-----VAGAINSG 1671

Qy 408 G-----NANKGVVIFTGASTVTDNIGNTPAVLAESVAGLILQIOGVVVKANAINLTD 459
Db 1672 GTASITAAARNATAADVNSAGDLTITATSGNLSA-GNVTTCQNLNESAGQAALSASQTEAG 1730

Qy 460 NASVVFTFGDSVTGSGIGTTELFATVNIAGITLRAGSLLAANNIDFGAASNLEFNGPAG 519
Db 1731 GNATLGGDMTLTGGI-----AAQNTG---TLTSGTTLTGSVAFGQQATINADG--- 1778

Qy 520 KNYNLIGTIA-NGNNATNLINAAAGTVIANDVSIQTV 554
Db 1779 -SINVSGSVAINGNLA-----ATATDSLSMGSV 1805

RESULT 2
US-10-282-122A-47372
; Sequence 47372, Application us/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
```

```
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47372
; LENGTH: 979
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47372
```

```
Query Match      12.3%; Score 373; DB 6; Length 979;
Best Local Similarity 26.2%; Pred. No. 2.9e-14;
Matches 174; Conservative 85; Mismatches 278; Indels 128; Gaps 29;

Qy 16 GLKTALEFTTSTAAILMTGSGVGLGAARTVTADGAELAAAGTNIQPGAGAFVAGSTLQYT 73
Db 341 GLNLLNNVTLGADLTVDGATGLALSGTIAGGALKASCTGTLTNGANSFTGGTNLTGGG 400

Qy 74 AFTV-TDADVSRALDLNFAAGLFSVTGDISLGSVDVTGGANKLAVNIDDLGLTLTLTGT 132
Db 401 GLVGTDTALGGGALNVSGAGGTGTSVGGTLLGN-----AINIDTGVTLGL--N 448

Qy 133 GTAAYGANPALLFOGQQAANNVTYAL-----GNITLGGANAGLTIASDPDVLGPTLT 185
Db 449 GSAGLSGALTAGNGLAQTFGTTLLGPNTFAGTTLSG---GGVAGTGAALG--TG 503

Qy 186 AGNIDG-GGIITDNTDRAINGTIGNTPAAOISI-GASTLSLGGAV-----IKATTTKLT 238
Db 504 ALHVTGTGSLATNAGGTLLGNVNLDAAGATLGLNGANGLSLGGTIAGNGGLAQSGTCTT 563

Qy 239 NAAPVLTITNANAVLTGAVDNTTGGD-DVGVNLNGALSQVTGNIGNTNSLATISVGAGT 297
Db 564 TLLGANTFSGGTALSGGGLIAGNCAALGTGALNVSGTGGTTLATNVGTVLGNVNLGAGA 623

Qy 298 A-----TLGGAVIKATTTKLTNAAVSLTTLTNAVLTGAVDNTTGG-----DNVG 340
Db 624 TLGVNGAVDLSLGAIGSGGQLAQTGAGTTLLGTNTFTGTTLSGGGLVAGNAGALGTG 683

Qy 341 VVNLSGALSQVTGNIGNTNSLATINIGAGVAT-LDGAIVIKATTTKLTDDASVLITFPV 399
Db 684 ALNVSGTGGTLATNVGTVLGNVNLGAGATLGLNGAA-----DLG 724

Qy 400 VTGAIDNTGN---ANKGVVIFTGASTVTDN---IGNTAVLAE-----VSVGAGLILQIOG--- 447
Db 725 LSGAISGAGGLALGCTGATTLTSGANTFTGGASLTGGGLIIVEGGSGALGIG-GALNVSGLG 783

Qy 448 -----GVVKANAINLTDNASVVTFTG--DSTVTSIGG-----TELFATV 485
Db 784 GSLSADLAAGAALANGINL--NGASLTINGANDLGLSAIGCTGGTLVLSGTTTSLGAN 841

Qy 486 NIGAITLRRAGGSAAANNIDFGAASNLEFNPGAKNYNLICTIANG---NNATNLINAA 542
Db 842 TFGGGTTLTLAGTLLAGSNTALGSGA-LNVAGTGG---ALGASVNGTALGNA-VNLGAGA 895

Qy 543 TVI---ANDVSI-----GTVAQINQNNKIFVINAKNADVDILDQAISFKGAASRLF 592
Db 896 TLALTGANDUGLNGVIAGSGGLAIGSGITTLTSGANTFSGGTTLSGGGLLAGNAGALGTG 955

Qy 593 LANVS 597
Db 956 ALNVT 960
```

```
RESULT 3
US-10-282-122A-47930
; Sequence 47930, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47930
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (101)..(101)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (236)..(236)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (254)..(254)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (256)..(256)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (268)..(268)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (282)..(282)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
```

```
; NAME/KEY: MISC_FEATURE
; LOCATION: (295)..(295)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (303)..(303)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (310)..(310)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (312)..(312)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (324)..(324)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (461)..(461)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (468)..(468)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (488)..(488)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (498)..(498)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (528)..(528)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (577)..(577)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (942)..(942)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (944)..(944)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (962)..(962)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (972)..(972)
; OTHER INFORMATION: X-any amino acid
; US-10-282-122A-47930
```

Query Match 12.1%; Score 366.5; DB 6; Length 1331;
Best Local Similarity 25.7%; Pred. No. 1.1e-13;
Matches 185; Conservative 65; Mismatches 235; Indels 235; Gaps 33;

QY 23 TTSTAAMLTGSG-----VLGAARTVADGAELAAGTNIGPGAGAFVAGSTLOYTGAF 75

Db 584 TLTGANTYTGTTINAGTLALGAGGSLASTGAMNLGAVGATDLGATGGQTLGALSGV 643

QY 76 TTTDADSVRLDLNNFAAGLFSVTGDISLGSVVDTGGANKLAVNIDDDGLTLTLTGTT- 134

Db 644 AGTNINLGNALTLGCTASCTFG-----GAIGGTG-----SLTLACTGTQ 683

```

Qy 135 -----AAY--CANPALLFOGQAAANNYYTALGNITLGG-----NAGLTIASDPDV--- 179
Db 684 TLNGASTYSGGTN---LNSGSGVGNNTALGTGTGVNNGGAATLDTNASTLVANAVNLGTG 740
Qy 180 -----LG---PTLAGNIDGGIITDNTDAANGTIGNT-NPAAOISIGASTLSLGGAVIK 231
Db 741 AALTLGGSNALDLSGATSGASLVKNGAATTLTGANTYTGCTTINAGTLAVGAGGSUAS 800
Qy 232 ATTKTLKLTDAASVILFTNANAVLT-GAVDNTTGGDVGVLNGLNG-----ALSQVTGNI 282
Db 801 TGAMNLASAGATLDSAAATGAQTILGALSGVAGTN---NLGGNALTILGGTASGTFSGAI 856
Qy 283 GNTNSLATISVAGTAVL-----GGAVIKATTKLTNAAASVILFTNNAV-LTGA----- 329
Db 857 GGTGSLTL--ASTGTQTLSGANTYTGTTIGGGSTLALGAGGSLASTGTVNLGAGATFD 914
Qy 330 VDNTTGGDVGWNLGSLAL-SQVT----- 352
Db 915 LRGASGAGTIGA--LTGAAGTNTVTSAAATGXRXAATAATRSAAASPAAXSPRAVRRX 972
Qy 353 -----GNIGNTNSLATINIGA-----GVATLDGAVI 378
Db 973 RARTRIPAARSAAARWRVAGGSLASTGAVTLAGTATLDSLGAATIGTLAGA-- 1030
Qy 379 KATTKLTDDASVILFTNPVVVTGAIDNTGN--ANKGVVIFTGASTVTDNIGNTAVL-- 433
Db 1031 GGTSVNLGANAALTINATNGTFFGAIGGTGGTVAGTGTQTLTGANTY---GGTTINGG 1087
Qy 434 AEVSAGLLOIQGVV-----KANAINLTDNASVVFT 467
Db 1088 STLALGAGGSLASGAVTLGAGATFDVSGATAQTGLTGTGAAGTNVNLGANAALTNGS 1147
Qy 468 GDSVTGSGIGT--ELFA-----TVNIGAGITLRAGGSLAANN----- 503
Db 1148 GNGTGGAGIGTGGVTFAGTGTQTLTGANTYTGTTINGGSLALGAGGSLASGSLNLA 1207
Qy 504 -----IDFGAASNL-----FNGPAGKNVNLGTTANGNATLNNAACTVANDVIGTV 554
Db 1208 GTGATFDMASAGAGTIGALSGVGTNVNL-----GAN-FTLTLNGS-----NNTFGGTI 1256

RESULT 4
US-10-282-122A-49854
; Sequence 49854, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 49854
; LENGTH: 2457
; TYPE: PRF
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49854

Query Match      12.08; Score 362; DB 6; Length 2457;
Best Local Similarity 27.6%; Pred. No. 5.3e-13;
Matches 195; Conservative 84; Mismatches 263; Indels 164; Gaps 42;

Qy 3 NISLKFQKATQKGLKTALETTSTAAM-LTGSGLGAARTVTAD-GAELAAAGTNIGPGA 60
Db 50 NVTVOAGQ-IVNAGSITAVQNLIAAAVOTLFGCTLAANGNVNASAGTTLNANTITAGR 108
Qy 61 GAFVAGSTLQVTGATVTDADVSRDLNPNFAAGLSVTCG-----ISLGSVVD-TGG-- 113
Db 109 QAAVSAATFDNCGSSASAD-OFTLSATNLVNH-GGSITQTGNGATSVNVSGLMDNTGTTI 166
Qy 114 --ANKLAVN---IDGLTLTLTGTT-----AAYGANPALLFQG----- 147
Db 167 QTNSTDLALGATLTLNNGKTIASSGSGTSLVKTCGLSNNGGTIATNGALMDGGAVSNRG 226
Qy 148 ----GQAAANNYYTALGNITLG--GA-----NAGLTIASDPDLGPTITLAGNI 189
Db 227 GTLAGOSSATLRLVSLDNSAGGYGAHRAVSVDITGLNAGGTIQAD-DALAVSAQSVTN 285
Qy 190 DGGIITDNTDA-AIN--GTGNTN-----PAAQISGASTLSLGGAVIKATT 234
Db 286 DGGTIANGGTGATTVNAAGALTNTNGLGNGNVSVSGASIDRSGGTITTAAGATTVOQS 345
Qy 235 TKLTNAAPVLT--NANAVLTGAVDNTTGGDVGVLN-----LNGALSOVGTGNTGT--NS 287
Db 346 STLGNRAGMIQCTGNVSASAGCAIDNTGGQIEVDGTNATMQLLAASLDNTNVRVANTNG 405
Qy 288 LATISVG-----AGTATLGG---AVIKATTTKLTAASVLT-----L 321
Db 406 ATTISAAITNSNTGGVAGAGAGAGGVDVTINATLTSNTNGAQLVGTGHDLTNLIAQFANN 465
Qy 322 TNAVLTGAVDNTTGGDVGVLNLSG-----ALSOVGTGNTGN-TNSLATNIG 367
Db 466 TNAILSGANNVTNLGNAAVINAGGSIHNGAIGLNVASLDNTTGTGNDAGSGGSAIA 525
Qy 368 AG-VATLDGAV-----IKATTTKLTDDASVILFTN-PVVVTG--AIDNTG--NANKGVVI 416
Db 526 TGLANOGGAIGSDQNLSTVTTNQLTGDRIIAGNDGAVTVNGDYTLDTGTQIQANHDLT- 584
Qy 417 FTGASTVTDNTGNTPAVIAEVSAGLLOIQGVVYKANAINLTNNAVTVFTGDSVTGSI 476
Db 585 FTTSNGFT-NOGTLGAVNALTVAANVNDNQAG-ADLNSNTTVNAA-----GSISNAGRI 637
Qy 477 GG-----TELFATVN-----IGAGITLRAG-----GSLANNIDFGAASNL-----BENGPA 519
Db 638 EGDVSTHSAALVNVATIVGNTVTLNAGSIANTGAAAA--FAATAVNLVSPGDISNTGG 695
Qy 520 KYNLIGTIANGNNATNLINA-----AGTVIANVDSIGTV-AQINTQ 560
Db 696 ANIFSLGDISIAADATRDANGLLANSRSVNDQS--TIEAQGNIE 739

RESULT 5
US-10-282-122A-47312
; Sequence 47312, Application US/10282122A

```



```

; GENERAL INFORMATION:
; APPLICANT: Wang, Liansu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47312
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47312

Query Match          11.7%; Score 353.5; DB 6; Length 827;
Best Local Similarity 26.3%; Pred. No. 3.4e-13;
Matches 186; Conservative 87; Mismatches 259; Indels 175; Gaps 36;

QY 24 TSTAATMLTGGG--VLGAART---VTADGAELAACTNIGPGAGAFVAGSTLQYTGAFTV 77
DB 5 SGAGGVQSGSGTTTLCGANTYGGGTTLSGGSLVSGSNTALGSGA-----LNVNGTIL 58
QY 78 TDADVSVRALDNNFAAGLFSVTGDISLGSVVDTTGGANKL---AVNIDGL-----TL 127
DB 59 S--ASVNGTTLGN-----AVT--LGAGATLGLNCANDLGLSGTISGSGLAQTGAGTT 107
QY 128 TLGTGCTAAYGANPALLFOGGQAANNY-----TALGNITLG-----G 166
DB 108 TLGTNTYDGGTT---LSGGGLVAGNGSALGTGALNVTGAGGSLGTSVGGTTLGNVNLG 164
QY 167 ANAGLTIASDPDVLGPITLGNIDGGIIT---TNTDAALNGTIGNTPAAQISIGAS 221
DB 165 AGATLVGAND-----LGLGGISGSGLSVSGPSTTTLTGNTYTGNTIG-----GGS 215
QY 222 TLSEL--GGAVIKATTTKLTNAAPVLTLTANAVLTGAVDNTTGGDDVGVNLNGLALSQVT 279
DB 216 TLAVGAGGSLSAGSALDLAGTGAALDISAATTPTQSGMLSGVAGTWN---NLGNTLTLA 272
QY 280 GNIGNTNSLATIS-----VGAGTATL-----GGAVIKATTTKL-----TNAASVLT 320
DB 273 GT-GNGNYAGTIGGTGGLTMAGTGTETLTGNTYTGATTINSGLTAIGAGGSLSASSPVN 331

```

```

QY 321 LTNAVLT-----CAVDNFTTGGDNVGV-----VNLSGALSQVVTGNIGNTNSLATINIGAGVAT 372
DB 332 LTGAGATFDVSGATTPTQTGTLTSLGVAGSTVNLGGN-NLTLTGGTNGTYGCTTAGAGGSLT 390
QY 373 LDGAVIKATT-----TKLTFDDASVLFTNPVVVVTGAIDNTGNANKGVIFTGASTVTT 424
DB 391 LSGTGTETLTGNNTVTGTTLSGGGTTLIAGSGSALGTGALNTSGAGT-----LAASPTG 445
QY 425 DNIGNTAVLAEYVSGAGLLQIOGGVVKANAIL-----TDNASVVTFTGDSVTT 473
DB 446 TTLGNA-----VNLGAGSTLTVGG---ANDLGLGGAISGGGLAVNGPSTTTLTGANTYTT 497
QY 474 GS--TGGTFLFATNIGAGITLRAGGSL-----AANNIDFGAASNLE-----FNGPGAKKNYN 523
DB 498 GSTTIGGG---STLAVGAGGSLSAGSAIDLSGTGATLDLSAATSPTSALSGGTGTNVN 554
QY 524 LIG---TIANGNNAT-----LNINAAGTVI-----ANDVSIGTVA---QIN 558
DB 555 LGGNTLTLLGGAGSGTYGGVIGGTGGLTLSGTGTETLTGNTYTGATTINSGLTAIGGNGS 614
QY 559 IONNKIFVINAKNADVDILDAQAISFKGAASRLFLANVLSLOMIELSL 605
DB 615 LSSSPVSLTAAGATLDLSGAASPOSTGAISGVAGSTVNLGNNTL 661

```

RESULT 6

```

US-10-369-493-10178
; Sequence 10178, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10178
; LENGTH: 4327
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(4327)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-10178

```

Query Match 11.6%; Score 350; DB 6; Length 4327;

Best Local Similarity 25.6%; Pred. No. 6.5e-12;

Matches 169; Conservative 90; Mismatches 236; Indels 166; Gaps 31;

```

QY 16 GLK--TALTTTSTAATMLTGGSVLGAARTV-----TADGAELAAAGTNIGPGAGAFV 64
DB 3061 GLKDGAAALAAANTNS-YITGSYVSVYDVTATMAQLSAMDTATTGTLTYGAGLKDSAAALV 3119
QY 65 AGSTLQYTGFTVTDADVS---VRALD-----LNNFAAGLFSVTGDISLGS---VVD 111
DB 3120 ANTNSVVTGAVTVTVTDAATTAQLGALDQDTGTVNYSKLAGIKDVTNSITDIDSGNVANA 3179
QY 112 GGANKLAVNIDGLTTLTGTGTAAYGANPALLFOGGQAAA-----NNYITALGNITLGG 166
DB 3180 GGA---TITVNDGIANLITDAGTVTGTNRNVTYTDAAASQAQLSQIDNYTTGALKYVTIKD 3236
QY 167 ANAGLTIASDPDVLGPITLGNIDGGIITDNTDAALNGT--IGNTNPAQIISIGASTLS 224
DB 3237 AVAALVA-----NTNSVYTSYAVSVTDAASMAQLSALDQD 3272

```

```

QY 225 LGGAVIKATTTKLTNAAPVLTILTNANAVLTGAVDNTTGGDDVGVNLNGALSOVVTGNTGN 284
Db 3273 TTGTL---TYKLTDAVANL-VTNTSVTGSV-NVTVSDIATISQLSSIDANTTGSVTY 3327
QY 285 T---NSLATISVGACT---ATLGGAVIKATT-TKLT-----NAAS 317
Db 3328 TOIGDAATLATNAGNYVKATIHVTVTDAAITIAQLTTIDGNTTGSLSVYTAGGVKDSAN 3387
QY 318 VLTTLNAVLTGAVDNTTGGDNVGVVNLNGALSOVVTGNTG-----NTNSLAT 363
Db 3388 LVVNTSVVTVGAV-NVSVTDTVSIQAQLSAVDEYTTGTLTYGAGVKDSVANLLVNTSVYT 3446
QY 364 -----INIGAVATLDGAVIKATT-----TKLTDASVLIFTNPVVVTGAI-----DNT 407
Db 3447 GSYAVSIITDVASMANL-SAIQDFTTGLTNYTKLSDTVSALVANTSVYTGSVNVVTITDNA 3505
QY 408 GNANKGVVIFTGASTVTDNIGNITAVLAESVSGAGLQIQGGVVKANAIN--LTDNASVVT 465
Db 3506 SWAN-----MSAIDQNTTGLTLYTKLSDTAALAANTSVYTGSVNVTVTDNATVAQ 3557
QY 466 FTG-DSTVTGSGIGTELPATVNIAGITPLRAGGSLAANNIDFGAASNLEFNGPAGKNY-- 522
Db 3558 LTTVDAATGTI---KVASVYDSSGNI-----SSNFAYVDGLGVSVYN 3597
QY 523 ---NLIG-TIANGNATLINAGTVIANDYSIGTVAQINTONKKIPVINAKNADVDILD 578
Db 3598 ANDNVMAVTAQDAIV-----TIADDDVLTITDSSNTQGSTFSALVTAGADATYLD 3651
QY 579 A 579
Db 3652 A 3652

```

RESULT 7

```

US-10-282-122A-49697
; Sequence 49697, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

```

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49697
; LENGTH: 3286
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
; US-10-282-122A-49697

```

```

Query Match 11.3%; Score 342.5; DB 6; Length 3286;
Best Local Similarity 25.1%; Pred. No. 1.2e-11;
Matches 176; Conservative 82; Mismatches 250; Indels 193; Gaps 34;
QY 23 TTSTAAILMTGSGVLGAARTVTADGAELAGNITGCPGAGAFV-----AGSTLQYT 72
Db 797 TLSGAATNVSGASVDNSNGTIEGDTLALSSGSLNRRGRLTQYGSADQTLISAGTLDNT 856
QY 73 GAFTVTADVSVRALDLNFAAGLFSVTGD-----ISLGVVVDGTGANKLA 118
Db 857 GGFVATNA-----TNLTVSQSVTNDTSGIQHAGSGTLALASAGTLLSNTGG----- 902
QY 119 VNIDDLGLTLTGTGTAAYGANPALLFQGOAA--ANNYTTALGNITLGGANAGLTASD 176
Db 903 -NVVTNGALDVGSTAVSNQGS-----LSAKGDATVRAOSLDNHAGSVVAGG--NLGANIA-- 954
QY 177 PDVLGPI-----TLAG---NIDGGIITDNDAAINGTI-CNT-----NPAA 214
Db 955 ---GALANQSGTSLGATTTVSGSSV--DNS---NCTIEGNTLALSSGSLNRRGRLT 1004
QY 215 QISIGASTLSGGAVIKATTTKLTNAAPVLTILTNANAVLTG-AVDNTGG---DDVGLVN 270
Db 1005 QYGSADQTLISAGGALDNTGTTATNA-----ANLTVSQSVTNDTSGIQHAGTGT 1056
QY 271 LN---GALSQVVTGNTGNTNSL-----ATISVGAGTATLGGAVIKATTTKLTNAASVL---T 320
Db 1057 VTPGALSVDVAGQIATNGALIARSASLDNSNGTVSAGSQAQVDANTSLTNRGGTLYKAG 1116
QY 321 LTNAVLTGAVDNTGG---DNVGVVNLGALSQVVTGNTGNTNSLATINIGAGVATLDGAV 377
Db 1117 LT-ATTQGAFTDQGSQVQTDGNSLVNAGGALSNTSGTI-----SVNGASGNAATATVSASS 1171
QY 378 IKATTTKLTDDASVLIFTNPVVVTGATDNTGNANKGVVIFTGASTVTDNIGNTAVLAESV 437
Db 1172 IDNTSGKLTNSGGA--TTTAAATGTTNSAGTMGGNGDVTILGAGTLLTNTASAKFVAARVA 1229
QY 438 -----VGAGLL-----QIQGGV-VKANAINLTDNASVVTET 467
Db 1230 SLNVNVRVNSGGTIYGGTALNLNQSGASVINDSGRLEGQDYSVRVASLANSAGALAN 1289
QY 468 GDSTVTGSGIGT-ELFATVNIAGI-----TLRAG 496
Db 1290 RDITASGVWSGDGTTMTAGRNGLAVTGDYTNAGSLHADGNMTVSATGTLTNTAGTLAAS 1349
QY 497 GSAAANNIDFGAASNLEFNGPAGKNYNICTIAN-----CNNATLN---LNAAGTVIAND 548
Db 1350 GALSAGANVNVNAGADINS-ASTTVNATGTLTNAGRIEGDSVTTISATLANTGTTLIGT 1408
QY 549 VSIQTVAQINIQNNKIFVINAKNADVDILDAQAIISFKGAAS 589
Db 1409 VQVNA--DVQNT-----GAAAIARAAQSILNIYASNS 1438

```

RESULT 8

```

US-10-282-122A-47709
; Sequence 47709, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```

```

Qy 325 -VUTG-----VDNTGGDNVGVVNLSGALSQVTGNIG----NNTSLATINI 366
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Db 945 VALSGAGTVIANGADCVLIDSVTGGGIA-----AGATSIAVGSGGVRLNGANATIAL 1000
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Qy 367 -GAGVAT---LDGAVIKATTKKLTDDASVLIFT-----NPVVYTGADNTGNANKGVI 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1001 SGTQVATTTGNGADGLTSTGAGARIATAATVVYRTACADARGVSVSGA--DSTLTANGTTIA 1059
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Qy 417 FTGAST---VTDNIGHTAVLAEYSVGAGLLQIQGGVVKANAINTLDNASVVYFTGDSTVT 473
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Db 1060 TTGANAHAIIMDG-GATALSGAKVTAS-GNAADGIVAQGGRIDATGSSLSASAAGSGAT 1117
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Qy 474 GSTGGTEFLTAVYNIGAITLRAGGSAAANNIDFGAASNLEFNPGPKNYNLICTIANGNN 533
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Db 1118 ANSGGV-LALT---GYALTGATAGVLTSDTLANGATSSVLDIGGSVTSATGPAPFAARGGT 1173
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Qy 534 ATLINAAGTVIANDVSYGTVQAIONKFIWNAKNADVILDAQAISFKGAASRFLF 593
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Db 1174 ADIAVRNGTVTAGN---GTLLNLANGSNVTESASAVNLAGDI-----VSDASSTGNVFL 1225
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Qy 594 ANVSLQMIELSLKIIPVLLTV 616
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |
Db 1226 ANET-----TLTKRIDPVALT 1242
      ||| :|| :|| | :|| | :|| | :|| | :|| | :|| | :|| |

RESULT 9
US-10-223-597-7
; Sequence 7, Application US/10223597
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: Nemellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDE
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/10/223,597
; PRIOR FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: US/09/379,931
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-10-223-597-7

Query Match          11.0%; Score 331; DB 6; Length 1026;
Best Local Similarity 27.3%; Pred No. 1,1e-11;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps

Qy 23 TTSTAAILMTGSGVYLGAART-----VTADG-----AELAGTNIGPGA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 TTTTGALAVTGGTAATVAQTAGNAVNTLTQADVTVTGNSSTTAVTVTQTAAT-----A 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 61 GAFVAGSTLOYTGAFVTYTDADV-----VRALDLNNEAAGL-----FSVTG-- 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 GATVAG---RVNGAIVITDSAAASATTAGKATVATLGSEGAATIDSSALTVVNLSGTGS 524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 102 -DISLGSVVDTTGANKLAVNIDBGLTILTTGTCTAAAYGANPALLFGQGAAANNFYTAG 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 LGIGRGALTATPATNTLTNVNG--LTTTGAIT-----DSEAAADDGFT--- 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 161 NITLGGNAGLUFTIAS-----DPDVLPGITLAGNIDGGG--LIIDNTDAAING-TIGNTPA 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 TINIAGSTASSTIASLVAADATTL-----NISGDARVTTISHTAAALGTITVTVNSVGA 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy 214 ---AOSIGASTLSLGA-----VIKATTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264
Db 620 TLGAEATG--LVFTGGAGROSLIGATTKAIVMGAGDDTIVTSSATL-GAGGSVNGGDG 676
Qy 265 -DVGVLNGLALSQVGTGNIGTNSLATISSVAGATATLG-----GAVIKATTK 311
Db 677 TDVLVANVNGSFSADPAFGGFETLRV-----AGAAQSHNANGFTALOLGATAGATT-- 730
Qy 312 LTNA-----SVLTJTNNAVLTGAVDNTTGGDNGVGVNLSGALSQVGTGNIGTNSLATINI 366
Db 731 FTNAVNVGLTVLAAPTCTTTVTLANATGTSDFENLTSLSSAALAAGTVA-LAGVETVNI 789
Qy 367 GA---GVATLDGAVIKATTKLTDDASVLFTFNVVVTGADNTGNANKGVVFTGAST 422
Db 790 AATDNTTAHDTLTQATSAR-----SIVVTG-----NAGLNL----- 823
Qy 423 VTDNIGNTAVLAESVSGAGLLQIOGVVVKANAINLTNDASVVVTFDGTSTVTSIGGTGTELF 482
Db 824 -TNTGNTA-----VTSFDSASVTGTAPATVFSANTTVGEV----- 858
Qy 483 ATVNIGAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGNNATL-NINA 540
Db 859 VTIRGAGADSLTSATANDTIIGGAGADTLVYTGTT-----DTFTGGTGADIFDINA 911
Qy 541 AGT-----VIANDVSTGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGA-ASRLFLAN 595
Db 912 IGTSTAFVTIIDAAG-----DKLDLVGISTNGAIDAGAFGA 949
Qy 596 VSL 598
Db 950 VTL 952

RESULT 10

US-10-282-122A-77944
; Sequence 77944, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77944
; LENGTH: 3705
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77944

Query Match 10.9% Score 329.5; DB 6; Length 3705;
Best Local Similarity 24.8%; Pred. No. 8.8e-11;

Matches 165; Conservative 88; Mismatches 233; Indels 179; Gaps 33;

Qy 23 TTSTAAILTSG-----VLGAARTVTADGAELAAGTNIIGP---GAGAFVAGST 68
Db 1144 TTVTAGELLGANGAFGQTSLLDIASGASANINGVSQVGVAVTNGVTITLCSGGVLTSSL 1203
Qy 69 LOYTGAFTVTDADVSRALDNNFAAGLFVSVDISLGSVVDTGANKLAV----- 119
Db 1204 LTNGGILDLTGG-----ALNLTAGGASTVAGGLTGAGTLNNGGN-LSVSAANSGLSG 1255
Qy 120 --NIDDDGLTLTGTGAAYGANPALLFQGGQAAANNYTTALGNITTLGGANAGLTIASDP 177
Db 1256 QTHIADVASVLTDTGTLGTS-----VEVLGTNLNNGANAMTNVLSG 1299
Qy 178 D-----VLGPITLAGNIDGGIITDNTDAAINGTIGNTNPAQAISIGASTLSGGAVIKAT 233
Db 1300 DGTINTNAAVTLSGNNSFSAHQITGDCEL--TVGQ---ASNLGASSATVNLG---TLT 1350
Qy 234 TTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGLALSQV-----TGN 281
Db 1351 SHLILNGVSEIANVLSGAVGTVDIIIGGADTALTANNNGFLGQYALAGNAGKLTVASTNN 1410
Qy 282 IGNTNSLA-----TISVGAGTATLGGAVIKATTTTKLTNAASVLTLT-----NAVLTGA 329
Db 1411 LGASSVALACAGDTLSLSGFNGTFGNSVGTSGVLQVTDDEAV-TLTSSNGCVSNAVITDI 1469
Qy 330 VDNTTGGDNGVNLGALSQVGTG---NIGHTNLSLATINIG-----AGVATL----- 373
Db 1470 ADATLNDLDDIALFN-----HVLTGNGLLNVAKNDASTAFDFGTVGGAFSGIVNLNTTF 1524
Qy 374 ---DGAIVKATTTKLTDDASVLFTFNPVVVTCADTNGNA---NKGVVIFTCAS---- 421
Db 1525 ALSADNAAARATLKLSDDS-----VTTVGATDRTLHGLDNGGTLTFDGSPPQS 1576
Qy 422 ---TVDNIGNTAVLAESVSGAG-----LLQTQGGVV---KANAINLTD 459
Db 1577 ANGVTVTDLALNSGTIS--ITGAGNWNENHPVTPPNVSLLEQDRGDLLELLEINAVTG 1634
Qy 460 NASVVTFTGSDTVTSGISGTELFATVNVIGAGITLRAGGSLAANNI-DFGAASNLEFNPA 518
Db 1635 NANNLDDLVDGTAITS--GTQ-----GVESAIOQGGSTVANAHHNYGLTSS---NGNG 1682
Qy 519 GK---NYNL-IGTIANGNNATLINAACTVIANDV-----SIGTVAQINIONNNKIFV 566
Db 1683 GSGLYVNTTSLADELLANGANALLATESG-LTANRVLNALNELFCVGGGLV-VDAQNGALT 1740
Qy 567 INAKN 571
Db 1741 ANGN 1745

RESULT 11

US-10-193-764-61
; Sequence 61, Application US/10193764
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Kleib, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764

```

; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 1180
; TYPE: prt
; ORGANTISM: Haemophilus influenzae
US-10-193-764-61

```

```

Query Match      10.7%; Score 322; DB 6; Length 1180;
Best Local Similarity 25.2%; Pred. No. 4.5e-11;
Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;

QY   48 AELAAGTNIGPGACAF-VAGSTLOYTGAFIV-----TDADVSRYALDLNNFAAGLFSVT 100
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   543 AEIOIGNISOKENLRFISSDKINITKOITIKGVNGENSDDSYKS-----QANLTFTT 596

QY   101 GDISLGSVVDPTGGANKLAVNIDDDLTLTLCTGCTAAYGANPALLF-----QGQA 150
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   597 KELKLTODLNISGFNKAKIVAKDSNLTIIGNSDDSGNTSAKTVTFNNVKDKSIADGHKV 656

QY   151 AANNYYATLG--NITLGGCA-NAGLTI-ASDPDVLGPITLAGNID---GGGIITDNDA 201
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   657 TLNKKVTKLSNDNNTTEGSDNNLTITAKDVEVNNNITSHTVTNNVAANGGITTKTGT 716

QY   202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTAAPVLTLTNANAVALTAGVDNTT 261
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   717 TINATAG-----VEITAHTGSIOGGI-----ESKP-----GSVTIVA 749

QY   262 GGDDVGVLNLNGALSQVTCGNIGNTNSLA-----TISVGAGIATLGGAV-----IKA 307
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   750 GGDTLAVGNISGNNAVTVANSGLATTLAGSTIKCETPISITTSQSNGICGSKGKTVNYKA 809

QY   308 TTTLKLTNAASVLLTLTNAVL-----TGAVDNTTGGDVGVNLSGALSQVTGNIGNTNSLA 362
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   810 TNSLTQTQADSKIATEGEANVTSKTSIIIGTISCGTVEVTEGLTTQAGSTITGESVT 869

QY   363 TI-----NITGAGVATLDGAVIKATTK-----LTDASVLIPT----NPVVVTGAINDTGNANK 412
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   870 TSQSGNIGT---GMISGGKVESATKDILITKGSSEIKATAGEVNVTSATGTTIDGTISGNT 926

QY   413 -GVWIFTCASTVDNIGNTAVTLAEVSVCGALLQIOGG--VVKANAINLTDNASVVFTFGD 469
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   927 VNVVTANTGDLVED-----AAKIDATGAATLTATSGKLTYKASSIITSANNOVNLAKD 981

QY   470 STVTSGIGGETELFATVNTIGAGTITIRAGGSLAANNIDFGA----ASNLEFNPGAGKNYMLI 525
    ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   982 GSIGGINNAANY--TLNNTTGALTITVKGSSINANNS---GTLVINAKDAELNGEASGNHTVV 1036

QY   536 GTIANGNATLINAAQGVIANDVDVSGHVAQIIONKKNIFVNAKNADVDILDQAQISFK 585
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   1037 -----NAT-NANGSGSVIAT-----TSSRVNITGCD-LITINGLN-----IISK 1072

QY   586 GAASRLFLFANYSLQMIELSLKIIPVLLTV 616

DB   1073 NGINTVLLKGKVID-----VKYIOPGIAVF 1097

```

```

RESULT 12
US-103-764-59
: Sequence 59, Application US/10193764
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-Ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: PROTECTIVE RECOMBINANT
: TITLE OF INVENTION: MOLECULAR WEIGHT PROTE
: FILE REFERENCE: 1038-139M1S
: CURRENT APPLICATION NUMBER: US/10/193,764
: CURRENT FILING DATE: 2002-07-12
: PRIOR APPLICATION NUMBER: 09/167,568

```

```

; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 1188
; TYPE: prt
; ORGANISM: Haemophilus influenzae
US-10-193-764-59

```

Query Match	10.7%	Score 322	DB 6	Length 1188
Best Local Similarity	25.2%	Pred. No. 4.6e-11		
Matches 159	Conservative	83	Mismatches 251	Indels 138
Gaps				
Qy	48	AELAAGTNIGPGAGF-VAGSTLQVTAFTV-----TDADVSVRALDNLNFAAGLFSVT	100	
Db	551	AEIQGNNISQEGNLTSSDPKINTKQITIKGVNGENSDDSYKS-----QANLTIKT	604	
Qy	101	GDISLGSVVDTGGANKLAVNIDDDGLTLTLTGCTGAAYGANPALLP-----QGGQA	150	
Db	605	KELKTODLNISGFNKAIVAKDSNLTIGNSDDSGNTSKTVTFNNVKDSKISADGHKV	664	
Qy	151	AANNTYTALG--NTILGGA--NAGLTI-ASDPDLGPITLAGNID--GGITLTDNDA	201	
Db	665	TLNSKVTKLSDNNTEGSSDNNTGLTITAKDVEVNNNITSHKTVNWSAANGITTKGT	724	
Qy	202	AINTGIGNTPAAQISIGASTLSLGGAVIKATTKLITNAAPVLTLTNANAVLTGAVDNTT	261	
Db	725	TINATAG-----VEITAHGTSIQGI-----ESKP-----GSVTIVA	757	
Qy	262	GGDDVGVNLNGLALSQVTCNTGNTNSLA-----TISVCAGATLGGAV-----IKA	307	
Db	758	GGDTLAVGNISGNNAVTVTANSGLTTLAGSTIKGTESITTSQSGNICGGKISGKTVNVYKA	817	
Qy	308	TTTKLTNAASVLTLTNAVL-----TGAVDNTTGGDNVGNVNLGSLASQVTGIGNTNSLA	362	
Db	818	TNSLTQADSKIEATEGEANVTSKTSIIGTISGCTVEVTEGLTTOAGSTIIGTESVT	877	
Qy	363	TI-----NTGAGVATLDGAVIKATTK--LTDASVLFT-----NPVVVTGAIDNTGNANK	412	
Db	878	TSSQSGNIG--GMSIGKVEYSAFKDLITKSGSEIKATAGEVNVTSATGIDTISGNT	934	
Qy	413	-GVVIFTGASTVTDNIGNTAVLAEVSGAGLLQIOGG--VVKANAINLTDNASVVTFGD	469	
Db	935	VNVVTANTGDLTVED-----AAKIDATGAATLTATSGKLTTKASSIITSANNQVNL	989	
Qy	470	STVTSGISGGTELFAPVNTGAGITLRAGSLAANNIDFGA----ASNLEFNGPAGKNYNLI	525	
Db	990	GSIGGNINAANY--TLNNTTGALTTVKGGSSINANS---GTLVINAKDAELNGEASGNHTV	1044	
Qy	526	GTIANGNATLINAAGVTVIANDVSIGVVAQIIONKNKFIYNAKNADVDILDAQAISFK	585	
Db	1045	-----NAT-NANGSGSVIAT-----TSSRNVTGCD-LITINGLN-----LISK	1080	
Qy	586	GAASRLFTANYSQWIELSLKIIPVLLTV	616	
Db	1081	NGINTVLLKGVKVID-----VKYIOPGTASV	1105	

RESULT 13
US-10-282-122A-49147
; Sequence 49147, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

```
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 49147
; LENGTH: 3073
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49147
```

```
Query Match 10.6%; Score 319.5; DB 6; Length 3073;
Best Local Similarity 25.0%; Pred. No. 2.7e-10;
Matches 175; Conservative 86; Mismatches 251; Indels 189; Gaps 35;

Qy 23 TTSTAAT---MLTG-SGVLGAAATVTDGA-----ELAA-----TNIGPGAGAFVAGST-- 68
Db 1031 TNGTDVAVNVLGSVTSITGGGAANPDGTIKKPSFAIGGTYTDVGSAINAAVSGGTAN 1090
Qy 69 -LQVTGAFVTDADVSVRALDNLNFAAGLFSVTDGDISLGSVVDTGGANKLAVNIDD---- 123
Db 1091 GVQY-----DTSAR-----TKVTLGCTGATTAVILSNVAN-GVANNDVAVNTQLQAM 1136
Qy 124 GLTLTLTGTAAYGA-----NPALFOGGQAAANNYITFALGNITLGGANAGLITIASD 176
Db 1137 CATICTSGVVTNSFVAYDDTTQGISLKGTTGTTITNVKAGALSASLDVAVNGSOLYQTN 1196
Qy 177 PDVLGPITLAGNI-----DGGGIITDNTDAAINGTIGNTNPAQAQISGASTLSL 225
Db 1197 ANV---ANVAGNVANVTNVNNTNNGGKIYFHA-----NSTLADSS-----ATGTSNVAI 1244
Qy 226 GGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDN-----TTGGDDVGVLNLCAL 275
Db 1245 GGAA-SATAANSVALGANSVAGRANAVSVGASGERQIINVANATNTGDVAVNLOOLAMG 1303
Qy 276 SOVGTGNTGNTNS-IATISVAGTATLGG-A-VIKATTTKLTNAASVLTITNAV-----L 326
Db 1304 ANVNSSGVVTNSFVAYDDTSKGVTLGSGGFTKAVT--LTNVANGVANADAVNNAQLKAM 1361
Qy 327 TGAVDNTTGGDNVGNVLSGALSQV--GNIGTNSLATINIGAGVATLDG---AVIKATT 382
Db 1362 GGTIDSSGNVTNFAVYDDTSKGVTLGAGSTKAVKALTNVANGVANADAVNNAQLKA-- 1419
Qy 383 TKLDDASVLFTNPVVVVTGAIDNTGNANKGVVIF--TGASTVT-DNIGNTAVLAESVVG 439
Db 1420 -----MGTTIDSSGNVTNFAVYDDTSKGVTLGAGSTKAVA----- 1457
Qy 440 AGLLIQGGVVVKANAINLT-----DNASVVT-----FTGDSVT--GSTGGTGF 482
Db 1458 --LTNVANGVANADAVNNAQLKAMGGTIDSSGNVTNFAVYDDTSKGVTLGAGSTKAV 1515
```

```
Qy 483 ATVNIGAG-----ITLRAGGSLAANN 503
Db 1516 ALTNVANGVANADAVNNAQLKAMGGTIDSSGNVTNFAVYDDTTLGKISLKGTTGTTIN 1575
Qy 504 IDFGAASNLFEENGPAKYNILIGTIANGNNATLNI-NAAGTV--IANDVSIGTVAAQINIQ 560
Db 1576 VKAGALSSTSLDAVNGS--QLYQTNANVANVANVAGNVANVTNVN-----NIT 1626
Qy 561 NN---KIFVINAKNADVDILDAQAISFKGAASRLFLANVSL 598
Db 1627 NGGKIKYFHANSTLADSSATGANSVATGGAASATAANSVAL 1667

RESULT 14
US-10-193-764-35
; Sequence 35, Application US/10193764
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 35
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-35
```

```
Query Match 10.5%; Score 316.5; DB 6; Length 915;
Best Local Similarity 23.8%; Pred. No. 6.6e-11;
Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

Qy 42 TVTADGAELAAAGTNIQPGAGAF-VAGSTLQVTGAFVTDADVSVRALDNLN-FAAGLFSV 99
Db 229 TNGDNTETIQIGGNISKEGNLTISSKVNITERITI-KAGVNGDSDSNATSNATITK 287
Qy 100 TGDISLGSVVDTGANKLAVNIDDLTLTGTCTAAYGANPALLFOGGQAAANNYITL 159
Db 288 TKELKLTNDLNLISGFNKAETITAKDNSNLT-----GDNSDAGNTDAK-- 329
Qy 160 GNITLGGANAGLTTASDPDLVGPITLAGNIDGGIITDNTDAAINGTIGNTNPAQAQISG 219
Db 330 -KVTFSNVVKDSKTSADHNV---TLNSKVETSG-DTSDSTEDGNNNTGLTITAKNVTN 383
Qy 220 ASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVLNLCALSOVT 279
Db 384 NNITSHKVTNLTAASENVTTKAGITINATTSVEVATKDGKIGIESNSGNVNTASGDT 443
Qy 280 GNIGN-TNSLATISVAG-TATLGGAVIKATTTKLTNAASVLTITNAV-----LTGAVD 331
Db 444 LNVSNITQNVTVAAASGAVTTTKGSTINAT---TCNANITTKTGEINGEVKSASGNV 499
Qy 332 NTTGGDNVGVNLSGALSQV--GNIGTNSLATINIGACVATLDCGAVIKATTTKLTDDASV 391
Db 500 ITASGNTLNVNITQNVTVTANS-----AITTEGSTINAT-----TGANI 543
Qy 392 LIETNPVVVVTGAIDNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVVGAGLIQIO--- 446
Db 544 -----TTQGNINGKVESSSGVTLTATQTLAVGNISGDTV--TITADKGLTKTTQSS 595
Qy 447 -----GGVVKANAINLTDNASVVTFTG-----STVTGSGIGG 478
Db 596 KINGTKSVTTSSSGDISGTISGNTSVSATGSLTTOAGSKIEAKTGEANVTSATGTIGG 655
Qy 479 TELFATVNI-----GAGITLRA--GSLAANNIDFGAASNLFEENFPAG 519
```

Db 656 TISGNTVNTANTDNLTKDARIKATGAVTLTATGGLTGTETSSDITSSNGQTLTA- 714
 QY 520 KYNLIGTIANGNNATLNINAAGTVIAN---DVSIGTVAQINQNNKIFVINAKNADVD- 575
 Db 715 KSSSIAGSI-NAANVTNLTTGTLTTVAGSKIEAASGTL-----VINAKDAQLDG 762
 QY 576 -----ILDAQAISFKGA--ASRLFLANVSLQMIELSLLKII 609
 Db 763 AASGDHTVVNATNANGSGSVIATTSRVNITGLDITINGLNII 805

RESULT 15
 US-10-193-764-37
 ; Sequence 37, Application US/10193764
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
 ; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
 ; FILE REFERENCE: 1038-1239MIS
 ; CURRENT APPLICATION NUMBER: US/10/193,764
 ; CURRENT FILING DATE: 2002-07-12
 ; PRIOR APPLICATION NUMBER: 09/167,568
 ; PRIOR FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 37
 ; LENGTH: 1222
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-193-764-37

Query Match 10.5%; Score 316.5; DB 6; Length 1222;
 Best Local Similarity 23.8%; Pred. No. 1e-10;
 Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

QY 42 TVTADGAELAAAGTNGIPGAGAF-VAGSTLOYTGAFVTDADVSVRALDINN-FAAGLFESV 99
 Db 536 TNGDNTETIQIGNISQKEGNLTISDDKYNITERII-KAGVNGDSDSNEATSANLTIK 594
 QY 100 TGDISLGSVVDTGGANKLAVNIDDLGLTLTGTGTAAAYGANPALLFQGGQAAANNTYTAL 159
 Db 595 TREKLTNDLINSFGNKAETAKDNSNLTJ-----GNSDAGNTDAK- 636
 QY 160 GNITLGGANAGLTIASDDPVLGPITLAGNIDGGIITDNTDRAINGTICNTNPAAQISIG 219
 Db 637 -KVTFSNVSKDISASDHNV---TLNSKVETSG-DTSDTEGNNNTGLTITAKNVTYN 690
 QY 220 ASTLSLGGAVIKATTKLTNAAPVLTLTNANAVLTGAVDNTGDDGVVLNMGALSQVT 279
 Db 691 NNITSHKTVNITASENVTTKAGTTINATGTSVEVTAKTGDIKGIENSGNVNITASGDT 750
 QY 280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTKLTNAAASVLTNNAV-----LTGAVD 331
 Db 751 LNVSNITGQNVTVAAASGAVTTTKGSTINAT-----TGNANITTKTGEINGEYKASGNNV 806
 QY 332 NITGGDNVGVNLSGALSQVTGNIGNTNSLATINIGACVATLDGAVIKATTKLTDDASV 391
 Db 807 ITASGNTLVNSNITGQNVTVTANS-----AITTEGSTINAT---TGDANI 850
 QY 392 LIETNPVWVTGADNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVSVAGLLOIO--- 446
 Db 851 -----TTQTGNGKNGKVESGVTLIATGQTLAVNIGSDTV--TITADKGLTTQTSS 902
 QY 447 -----GGVVKANAINLTDNASVWFTFGD-----STVTGSIG 478
 Db 903 KINGTKSVITSSQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVTSATGTIG 962
 QY 479 TELFATVNI-----GAGITLRA--GSLAANNIDFGAASNFENGPAG 519
 Db 963 TISGNTVNTANTDNLTKDARIKATGAVTLTATGGLTGTETSSDITSSNGQTLTA- 1021

QY 520 KYNLIGTIANGNNATLNINAAGTVIAN---DVSIGTVAQINQNNKIFVINAKNADVD- 575
 Db 1022 KSSSIAGSI-NAANVTNLTTGTLTTVAGSKIEAASGTL-----VINAKDAQLDG 1069
 QY 576 -----ILDAQAISFKGA--ASRLFLANVSLQMIELSLLKII 609
 Db 1070 AASGDHTVVNATNANGSGSVIATTSRVNITGLDITINGLNII 1112

Search completed: April 14, 2003, 16:16:10
 Job time : 96 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:07:44 ; Search time 21 Seconds
(without alignments)
2833.677 Million cell updates/sec

Title: US-09-800-065-2
Perfect score: 3022
Sequence: 1 MANISLKLQRAIQKGLKTA.....MIELSLKLIYVLLTVVVS 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:* *
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1510	50.0	2021	2 A97859	190-KDa cell surfa
2	1494.5	49.5	2249	2 A41477	190K surface antig
3	1103	36.5	1029	2 T30852	outer membrane pro
4	452.5	15.0	2340	2 B71704	cell surface antig
5	372.5	12.3	1651	2 JC1340	outer membrane pro
6	333.5	11.0	1026	2 A48995	paracrystalline su
7	333.5	11.0	1073	2 C87374	S-layer protein Rs
8	333	11.0	1643	2 D71630	outer membrane pro
9	331.5	11.0	1655	2 E97835	hypothetical prote
10	329.5	10.9	3705	2 AD0123	probable autotrans
11	329	10.9	2554	2 AB3528	extracellular seri
12	323.5	10.7	1052	2 AF2959	conserved hypothet
13	323.5	10.7	1341	2 H98323	hypothetical prote
14	310.5	10.3	652	2 E97857	cell surface antig
15	308.5	10.2	256	2 G97778	cell surface antig
16	302	10.0	1145	2 G87284	hypothetical prote
17	299.5	9.9	1645	2 JN0896	crystalline surfac
18	298	9.9	3013	2 AB0480	probable invasiv y
19	295.5	9.8	1369	2 T17504	hypothetical prote
20	293.5	9.7	1910	2 AF0394	probable adhesin h
21	286	9.5	820	2 T17519	cell surface antig
22	286	9.5	1477	2 B43855	high-molecular-wei
23	282.5	9.3	1335	2 T17508	glycoprotein Vp260
24	281.5	9.3	1109	2 A56143	surface-array prot
25	278	9.2	591	2 AC3528	extracellular seri
26	277	9.2	5188	2 B85547	probable RTX fami
27	277	9.2	5291	2 F90696	hypothetical prote
28	276.5	9.1	1300	2 S07575	outer membrane pro
29	276	9.1	1407	2 B72078	polymorphic outer

30	273.5	9.1	1536	2 A43855	high-molecular-wei
31	273	9.0	1749	2 S75138	hypothetical prote
32	270.5	9.0	1176	2 T18042	ice nucleation pro
33	270	8.9	962	2 AG2444	hypothetical prote
34	270	8.9	1158	2 AF1852	hypothetical prote
35	269.5	8.9	1280	2 AB1981	hypothetical prote
36	269	8.9	1461	2 E90696	hypothetical prote
37	269	8.9	1461	2 A85547	hypothetical prote
38	267.5	8.9	5627	2 C83339	hypothetical prote
39	266	8.8	918	2 AB2445	hypothetical prote
40	266	8.8	1428	2 AC2224	hypothetical prote
41	264.5	8.8	1070	2 A10484	probable autotrans
42	263.5	8.7	3535	2 E83641	probable hemagglut
43	263	8.7	1649	2 C86822	hypothetical prote
44	262.5	8.7	1268	2 B99789	hemagglutinin/hemo
45	262.5	8.7	1270	2 E85649	hypothetical prote

ALIGNMENTS

RESULT 1

A97859

190-KDa cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: A97859
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.,
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97859
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03811.1; PID:g15620410; GSPDB:GM00173
C:Genetics:
A:Gene: rompa

Query Match 50.0%; Score 1510; DB 2; Length 2021;
Best Local Similarity 37.3%; Pred. No. 6.4e-67;
Matches 402; Conservative 53; Mismatches 127; Indels 496; Gaps 18;

QY	1	MANISLKLQKALQGLKALFTTSTAALMTGSGVLG-AARTVTADGAELAGTNI	GP 59
DB	1	MANISPKLFQKALQGLKALFTTSTAALMTSSGALGIAVSGVIATNNNAFSDNV	GN 60
QY	60	---AGAFVAGST-----LQYTGAFVTVDADVSRALDINFA---AGLFSVTG	101
DB	61	WNEITAGVANGTPARGPQNNAFTYGGDYIT-ADVADHIITAINVADTPIGL-NIAQ	118
QY	102	DISLGSVDTGGANKLAVNIDDLTLTGTGTAAAGANPALLFQGGQAAANTYALGN	161
DB	119	NTVGSIV-TGG-NLLPVTITAGKSLTLNGNADA--AN-----HGFAPADN-YTGLN	168
QY	162	ITLGGANAGITIASDPLGPITLAGNIDGGIITDNTDAIINGTGTNPAQISIGAS	221
DB	169	IALGGANAALIIQSAP--AKITLAGNINGGGIITVKTDAIINGTGTNPAQISIGAS	226
QY	222	TLISLGGAVIKATTTKLTNAAPVLTLNANAVLTGAVDNTTGGDDVGLNINLNGALSQV	281
DB	227	IATLEGAAIKATTTKLTNAASVLTNLNNAVLGTADNTTGVNDVGNLNGALSQV	286
QY	282	IGNTNSLATISVGAGTATLG-----	301
DB	287	IGNTNALATISVGAGKATLGGAVIKATTTKLTDNASAVTFTNPVVTGATDNTGNANNGI	346
QY	302	-----	301
DB	347	VFTTGDSTVTGNIGNTNALATISVGAGKATLGGAIKATTTKLTDNASAVTFTNPVVTG	406
QY	302	-----	301

Db 407 AIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGGAIKATTTKLTNDAS 466
Qy 302 -----
Db 467 AVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGGAI 526
Qy 302 -----
Db 527 KATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISV 586
Qy 302 -----
Db 587 GAGKATLGGAIKATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNT 646
Qy 302 -----GAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTTGGDNV 339
Db 647 IGNTNALATVNVGAGIATLEGAVIKATTTKLTNAASVLTLTNVNAVLTGAIDNTTGV 706
Qy 340 GVNLSGALSOVGTGNTGNNTSINATINIGAGVATLDGAVIKATTTKLTDDASVLTFTNP 399
Db 707 GVLNLSGALSOVGTGNTGNNTSINATINIGAGVATLDGAVIKATTTKLTDDASVLTFTNP 766
Qy 400 VTGAIDNTGNANNGIVFTGDSVTGNTGNNTSINATINIGAGVATLDGAVIKATTTKLT 459
Db 767 VTGAIDNTGNANNGIVFTGDSVTGNTGNNTSINATINIGAGVATLDGAVIKATTTKLT 826
Qy 460 NAS-----
Db 827 NASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATL 886
Qy 463 -----VVTFTGDSVTGSIIGTTELFAT 484
Db 887 ALIKATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNAL 946
Qy 485 VNIGAGITLRAGSLAANNIDFGAASNLEFNGP--AGKN---YNLIGTIANGNNATLNIN 539
Db 947 VNVGAGVTLQAGSLDANNIDFGAASNLEFNGP--AGKN---YNLIGTIANGNNATLNIN 1006
Qy 540 AAGTVIANDVSGITVAQINONKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
Db 1007 TK-LLTAYHETIGTVAEINIGAGNLTIDASAGDVTILNAQDIIHFRALDSALVLSNLT 1063
RESULT 2
A41477
190K surface antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
C:Accession: A41477
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
Infect. Immun. 58, 2760-2769, 1990
A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated, ne
A:Reference number: A41477; MUID:90354033; PMID:2117568
A:Accession: A41477
A:Molecule type: DNA
A:Residues: 1-2249 <AND>
A:Cross-references: CB:W31227; NID:g152465; PIDN:AAA26380.1; PID:g152466
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for residue
C:Keywords: surface antigen; tandem repeat
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>
Query Match 49.58; Score 1494.5; DB 2; Length 2249;
Best Local Similarity 51.18; Pred. No. 4.2e-66;
Matches 364; Conservative 58; Mismatches 123; Indels 167; Gaps 18;
Qy 20 ALFTTSTAAILTGS-----GVL---GAARTVTAD--GAELAAGTNGIPG---A 60
Db 613 SVLTLTNANAVLTGAIDNTGCDNVGVNLNGLSALSOVGTGNTGNNTSINATISVGAGTATLG 672
Qy 61 GAFVAGSTLQYTGAFV---TDADVSVRALD--LNNFAAGLFSVTGDISLGSVVDVTGGANK 116
Db 673 GAVIKATTTKLTNAASVAVKFTNPVVVTGAIDSTGNANNGIVFTGDSVTGNTGNNTNAL 730

Qy 117 LA-VNIDDLGLTTLTGCTTAAYGANPALLFOGGGAAAANNVTALGN-----ITLGGANAGL 171
Db 731 LATNV-----GAGTATLGG-----AVIKATTTKLTNAASVLTLTNANAVL 771
Qy 172 TIAD-----PDVLGPITLAGNIDGGIITDNTDRAINGTNGTNPAAQISIGASTLSLGG 227
Db 772 TGAIDNTTGGDNVGVNLNGLA-----SOVTDIGNTSINATISVGAGTATLGG 820
Qy 228 AVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDGVVLNGLSALSOVGTGNTGN 287
Db 821 AVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDNVGVNLNGLSALSOVGTGNTGN 880
Qy 288 LATSIVGAGTATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTTGGDNVGVNLS 345
Db 881 LATSIVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLN 940
Qy 346 GALSQVTGNTSINATINIGAGVATLDGAVIKATTTKLTDDASVLTFTNPVVVTGAID 405
Db 941 GALSQVTGNTSINATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVVVTGAID 1000
Qy 406 NTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAGLLQIOGGVVKANAILTNDAS--- 462
Db 1001 NTGNANNGIVFTGDSVTGNTGNNTSINATINIGAGVATLDGAVIKATTTKLTNDASAVT 1060
Qy 463 -----
Db 1061 FTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTSINATINIGAGVATLGGVVKAN 1120
Qy 463 -----VVTFTGDSVTGSIIGTTELFATVNIAG 490
Db 1121 TINLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTSINATINIGAG 1180
Qy 491 ITRAGSLAANNIDFGAASNLEFNGP--AGK--NYNLIGTIANGNNATLNINAAFTVI 545
Db 1181 ITRAGSLAANNIDFGAASNLEFNGP--AGK--NYNLIGTIANGNNATLNINAAFTVI 1239
Qy 546 ANDVSGITVAQINONKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
Db 1240 ASHUTIGTVAEINIGAGNLTIDASAGDVTILNAQDIIHFRALDSALVLSNLT 1291
RESULT 3
T30852
outer membrane protein A - Rickettsia conorii (fragment)
C:Species: Rickettsia conorii
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30852
R:Gilmore Jr., R.D.
Gene 125, 97-102, 1993
A:Title: Comparison of the ompA gene repeat regions of Rickettsiae reveals species-s
A:Reference number: Z20904; MUID:93194085; PMID:7680636
A:Accession: T30852
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1029 <GIL>
A:Cross-references: EMBL:L01462; NID:g152485; PID:g152486; PIDN:AAA99908.1
C:Genetics:
A:Gene: ompA
Query Match 36.58; Score 1103; DB 2; Length 1029;
Best Local Similarity 42.38; Pred. No. 3.4e-47;
Matches 283; Conservative 60; Mismatches 194; Indels 132; Gaps 18;
Qy 16 GLKTALETTSTAAILT--GSGVLGAARTVTADGAELAAGTNGIPGAGAF-----VAGSTL 69
Db 365 GNTNALATISVGAGKATLGGAIKATTTKLTNDASQVFTNPVVVTGAIDNTGNANNGIV 424
Qy 70 QYGAFTVTDADVSVRALD--LNNFAAGLFSVTG-----ISLGS 107
Db 425 TFGTGGSTVGTGNTGNNTSINATINIGAGVATLDGAVIKATTTKLTNDASQVFTNPVVVTGA 484
Qy 108 VVDVTGGANKLAVNIDDLGLT-----LTGTGAAYG----- 138

Db 485 IDNTGNANNGIVTFTGDSVTGNTGNATATISVGAGKATLGGAIIKATTTKLTNDASQ 544
QY 139 ---ANPALL-----FOGQAAANN--TYTALGNTILGANAGL-----171
Db 545 VTFTNPVVVTGAIDNTGNANNGIVTFTGDSVTGNTGNATATNVG---AGLLRVOGG 601
QY 172 -----TI-----ASDPDLVGPITLAGNID-----GGGIITDNTDAAINGTICNTNPAAQI 216
Db 602 VYKSNITNLTDNASQVTFNPVVPVVTGAIDNTGNANNGIVTFTGDSVTGNTGNATATV 561
QY 217 STGASTLSGGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVLNLNGALS 276
Db 662 NVGAGLLRVOGGVKSNTINLTNDASQVTFNP-VVVTGAIDN-TGNANNGIVTFTGN-S 718
QY 277 QYTGNTGNATATISVGAGTATLGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTT 334
Db 719 TVTGNIGNTNALATNVNGAGIATLGAVIKATTTKLTNAASVLTLTNNAVLTGAIDNTT 778
QY 335 GGDNVGVNLGALSQVTVTGNIGNTSLATINIGAGVATLDGAVIKATTTKLTDDASVLIF 394
Db 779 GVDNVGVNLGALSQVTVTGNIGNTNALATISVGAGKATLGGAIVIKATTTKLTNDASQVTF 838
QY 395 TNPVVVTGAIDNTGNANNGVIFTGASTVTDNIGNTAVLAESVSGAGLLQIOGGVVKANA 454
Db 839 TNPVVVTGAIDNTGNANNGIATFTGDSVTGNTGNATATNVGAGLLRVOGGVVKSN 898
QY 455 INLTNDAS-----VVFTGDSVTGSGTGTGTELFATVNIAGI 491
Db 899 INLTNDASAVTFTNPVVTGAIDNTGNANNGIVTFTGDSVTGNTGNATATISVGAGK 958
QY 492 TLRAGGSLAANNIDF-GAASNLEFNGPAGKNYLNIGTIANGNNATLINAAQTVIANDVS 550
Db 959 AILGGAIKATTTKLTNDASQVTFNPVVT-CAIDNTGNANNGIVTFTGDSVTGNTGN 1017
QY 551 IGTVAQINI 559
Db 1018 TNALATNV 1026

RESULT 4

B71704

cell surface antigen (sca3) RP451 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000

C:Accession: B71704

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, D.

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499; PMID:9823893

A:Accession: B71704

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2340 <AND>

A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAAL4908.1; PID:g386100

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: sca3; RP451

Query Match

Best Local Similarity 15.0%; Score 452.5; DB 2; Length 2340;

Matches 194; Conservative 84; Mismatches 232; Indels 171; Gaps 34;

QY 26 TAAIMLTGSGVLGAATVTDAGLAAGTNGICGAGAFVAGSFTQYTGAVTVD---ADV 82

Db 407 TDCVTNNNVNGGTLIIINAKNTISAKLLNATKAKTOIANLTMNPSAGDISDIRAON 466

QY 83 SVRALDNFAGLSFVTDGDISL---GSVVDTGKANKLAVNIDGLTLTLTGTAAYGA 139

Db 467 TIYIDAKN-----GNVLLNNNAKIIPEGADSMALINTGYTADRTFI-----510

QY 140 NPALLFOGQAAANNYYTALGNTILGGANAGLTIASDPVLGPTIAGN-----ID 190

Db 511 -----IYNLNQSGNDEY---GIVKBAIKKVITIANQS---GPYTTIGQDNTHLRLKELIVE 560

QY 191 GGG-IITDNTDAAINGTICNTNPAAQISIGASTLSLGA--VIKATTTKLTNAAPVLT 246
Db 561 GAGDIIIDT-----IFTKLLSINSTGTQTFN-RTLDLGAAGNTAFKKGHTLVVNG-----610
QY 247 TNANAVLTGAVDNTTGGDDVGVLNLNGALSQVTVTGNIG-NTNSLATISVGAGTATLGAVI 305
Db 611 -----VTGSI--TTSENNQGITLINS--GNITGVIGTNELGLKLVNIGADPVTCSANVF 660
QY 306 KATTTKLTNAASVLTLTNAV-LTGAVDNTTGGDNVGVNLGALSQVTVTGNIG-NTNSLAT 363
Db 661 --ASVALTPSPSVLILADGVTLTGEV--TTHNNTKGVLSL-GTGSNITGOIGTNSAALEK 715
QY 364 INIGAGVATLDGAVIKATTTKLTDDASVLIFTNPVVVTGAIDNTGNANNGVIFTGASTV 423
Db 716 INIGAGASNID-SNIYAGSVTLTDQTSELTSFLANDVVVNSNIITTAGNNSGKLIFTGNGOI 774
QY 424 TDNIG-NTAVLAEV-----GITLRAGGSLAANNIDFGA-ASNLEFNGPAGKN 521
Db 775 TGNIGANGAALOEVEVFNGTNTNIGGTANSQNTVAHSAANVITGLTTGALKYKDTGTIIA 834
QY 446 QGVV-----KANAINLTDNASV-----VTFTGDSVTGSGIG--GTFLFA 483
Db 835 HGGLVGDIDFNKAGKFFILGDGAMIDGSLCNGGVAGTLDFFIGDGNVTQNIQADNANSIS 894
QY 484 TVNIGA-----GITLRAGGSLAANNIDFGA-ASNLEFNGPAGKN 521
Db 895 TINIQDNTKNTVIANDIFVDNIHFTNGGILQLOGGLTTHNIDFGANGTLEFNG--NNT 952
QY 522 YNLIGTIANGNNATLINAAQTVIANDVSIGTVAQINI-----QNNKIFVNAKNADVD 575
Db 953 YNLNAILVNGQGL--NAFTNLKASDDDTGIVKIINIQTGPQN---FTIQVNNKNLT 1007
QY 576 ILDA-QALSFKAASRLFLA 594
Db 1008 LVSSVSSINFGDANSQILS 1028

RESULT 5

JCI340

outer membrane protein B precursor - Rickettsia japonica

C:Species: Rickettsia japonica

C>Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 13-Nov-1998

C:Accession: JCI340

R:Yan, Y.; Uchida, T.

Chinese J. Microbiol. Immunol. 16, 220-226, 1996

A:Title: Cloning and sequence analysis of the gene encoding the 120kDa outer membrane

A:Reference number: JCI340

A:Accession: JCI340

A:Molecule type: DNA

A:Residues: 1-1651 <YAN>

A>Note: the authors translated the codon GGT for residue 241 as Cys, CAA for residue

C:Genetics:

A:Gene: ompB

C:Keywords: membrane protein

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1651/Product: outer membrane protein B #status predicted <MAT>

Query Match

Best Local Similarity 12.3%; Score 372.5; DB 2; Length 1651;

Matches 185; Conservative 103; Mismatches 260; Indels 213; Gaps 39;

QY 13 IQKGLKTAFTTSTAAMLTGSG-VLGAA---RT-----VTADGA---ELAAGTNIQPG 59

Db 8 LKRLISAGLVTAFTAFIVASFGSANGAALQQNRTTKVAATTVGVGFDOTAPANVAVA 67

QY 60 AGAFVAGSLOXTGAVTVDADVSVALDINFPAGLSFVTDISLSGVVD-TGGANKLA 118

Db 68 PNAVI---TANANNGLNLTNPAGSFNGLFLSNANNLAVTVSEDTTLGFINNAANNANCFN 124

QY 119 VNTDDGLTTLTGTG-----TAAAGA-NPALLFOGQAAANNYYTALGNTILGGANAGL 171

Db 125 LTLNAGKTLTITGOGITNVQSAATHNAQNIQVAKFNFGAAIANNDLSGLGTIDFGAAASTL 184

Qy 172 T---IASDPDVLGPITLACNIDGGGIITDNTDAAINGTIGNTPAAQIS-----I 218
 Db 185 VFDLANPTTQKAPLILADN-----ALIVNGA-----NGTLNVTNGFIQVSDKSFATVKAINI 236
 Qy 219 G-----ASTLSIGGAVIKATTTKLTNAAPVLTLNA---NAVLTGAVDNTT 261
 Db 237 GDQGFMENTNANALNLAQGGTTINFNGTGTGRLVLILSLKNGAATNFNITGSL----- 292
 Qy 262 GDDVGLNLGAL--SOVTGNIGNTSLATISVGAG-----TATLGG----- 302
 Db 293 GGNLKGIELTVAINGOLIANAGANAVICTNNGCAGAGFVSVVNDNGKAATIDGQVYA 352
 Qy 303 --AVIKAT-----TTKLTNAASVLTLTNAVLTGAVD----- 331
 Db 353 KDWIQSANANGOVNFRHIVDVGIDGTTAFKTAASIVAITQNSNFGTDFGNLAAQVTP 412
 Qy 332 -----NTTG--GDNVGV--NLSSGALSOVT--GNIGNTSLATINI--GAGVATLD 374
 Db 413 DTMILTGNFTGDANNPGNTAGVIITFDANGTLESADANAVTNNTITAIESAGGAVWOLS 472
 Qy 375 GAVIKATTTKLTDDASVLIFTPVVTGAIDNT-----GNANKGVVIFTGASTVTDNIGNT 430
 Db 473 GT--HTAELRLGNAGSVFKLADGVINKVQTVLVGCVLAAGAITLDGSATITCDIGNA 530
 Qy 431 ---AVLAESV-----GAGLL-----QTQGVVKANAIN----- 456
 Db 531 GGGALQISITLANDATKTLTLAGANIISANRGTFINFGANGGTIKLTSTQNNIVVDCDLAI 590
 Qy 457 LTD-----NASVVFITGDSVTGSG--GTELFATVNICAGITLRAGGSLAANNIDFG 507
 Db 591 ATDQGVVDASSLNAQTLTISGTIGIIGANNQTLGOFNIGSKTTLNSGNVAINELVIG 650
 Qy 508 AASNLEF-----NGPAGKNYNLIGTIANGNATLNNAAGTVIANDVSICTVAOI 557
 Db 651 NNGSVQFAHNTYLITRTNNAAGQKILFNPVN--NNTL-----AAGTNLGS--ATNPLAEI 704
 Qy 558 NIQNNKIFVINAKNADVD--ILDAQISEFKGAASRLFLANVS 597
 Db 705 N-----FGSKGARLDTVLNV-----GEGVNLATNIT 731

RESULT 6

A48995

paracrystalline surface layer protein RsaA - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000.

C;Accession: A48995

R;Gilchrist, A.; Fisher, J.A.; Smit, J.

Can. J. Microbiol. 39, 193-202, 1992

A;Title: Nucleotide sequence analysis of the gene encoding the Caulobacter crescentus pa

A;Reference number: A48995; MUID:93007489; PMID:1393820

A;Accession: A48995

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-1026 <GIL>

A;Cross-references: GB:AF062345; GB:M22663; GB:M84760; NID:96064104; PIDN:AAC38665.2; PI

A;Experimental source: CB15A, ATCC 19089

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIN:116173, NCBIPI:116174)

Query Match 11.0%; Score 333.5; DB 2; Length 1026;

Best Local Similarity 24.7%; Pred. No. 2.7e-09;

Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

Qy 16 GLKTALETTSTAALMLTSGSVL-----GAARTVTAD-GAELAAGT-----NIGP 58

Db 315 GIETMNTVTSAAITLNTSSGVTGLTALTNTWTSAGAAQVTTAGACONLTATTAAQAANNVAV 374

Qy 59 GAGAFV-----AGSTLQYTGAVTVDADVSRYALDLNNEAGLESFVTCIDSL----- 105

Db 375 DGGANVTAVSTGVTSGTTT--VGANSAASGTVSVSVANSSTTTTGTGAVTGGTAVTVAQT 432

Qy 106 -GSVVD-----TGCANKLAVNIDDLTLTLTGTTTAAAYGANPALLFGGQAAANN 154
 Db 433 AGNAVNTTLTQADVTTCNSSTTAVTVTQ--TAAATAGATVAGRVNGAVTITDSAAASAT 490
 Qy 155 TTAALGNITLGGANAGLTIASDPDVLGPITLAG-----NIDG 191
 Db 491 TAGKTATVTLGSGFAA--TI--DSSALTTVNLSSGTGTSIGIGRGALTATPTANTLTNLVNG 547
 Qy 192 --GGLINDNTDAA-----INGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNA 240
 Db 548 LTTTGAITDSEAAADGGFTTINIAGSTASSIASLVAADATTLNLSGDARVTIISHTAA 607
 Qy 241 APVLTTLNA-----NAVLTCAVDNT----- 260
 Db 608 LTGITVTSVSGATLGAELATGLVFTGGAGADSIILLGATTKAIVMGAGDUTVTVSSATLGA 667
 Qy 261 ---TGGD--DVGVLNLNGALSOVTGNIGNTNSLATISVGAGATATLG-----G 302
 Db 668 GGSVNGGDDTDLVANVNGSFSADPAFGGFETLRV-----AGAAAQGSNANGFTALQLG 723
 Qy 303 AVIKATTTKLTNA-----SVLTTLTNAVLTCAVDNTTGGDNVGVVNLSSGALSOVTGNIGN 357
 Db 724 ATAGATT--FTNAVNVGLTVLAAPTGTTTVTLANATGTSDFNLTLSAALAAAGTVA-- 780
 Qy 358 TNSLATINIGA-----GVATLDGAVIKATTTKLTDDASVLIFTPNVPVVTGAIDNTGNANKG 413
 Db 781 LAGVETVNIATDTNTTAHVDTLTLOATSAK-----SIVVTG-----NAG 820
 Qy 414 VVIFTGASTVTDNIGNTAVLAESVVGAGLLQIQGVVKANAINLTDNASVVTFTGDSVT 473
 Db 821 LNL-----TNTGNTA-----VTSFDASAVTGTGSAVTFVSANTTV 855
 Qy 474 GSIIGTELFATVNICAGITLRAGGSLAANNIDFGA--ASNLEFENGPAKNYNLIGTIANGN 532
 Db 856 GEV-----VTIRGAGADSLTGSATANDTIIGAGADTLVYTGCT-----DTFTGGT 902
 Qy 533 NATL-NINAGT-----VIANDVSICTVAQINQNNKIFVINAKNADVDILDAQAISEFKA 587
 Db 903 GADIFDINAIGTSTAFVITTDAAVG-----DKLDLVGISTNGA 940
 Qy 588 -ASRLFLANVSL 598
 Db 941 IADGAFGAAVTL 952

RESULT 7

C87374

S-layer protein RsaA [Imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: C87374

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

n, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: C87374

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1073 <STO>

A;Cross-references: GB:AE005673; NID:gl3422297; PIDN:AAK22991.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC1007

Query Match 11.0%; Score 333.5; DB 2; Length 1073;

Best Local Similarity 24.7%; Pred. No. 2.8e-09;

Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

Qy 16 GLKTALETTSTAALMLTSGSVL-----GAARTVTAD-GAELAAGT-----NIGP 58

Db 362 GIETMNTVTSAAITLNTSSGVTGLTALTNTWTSAGAAQVTTAGACONLTATTAAQAANNVAV 421

RESULT 9

E97835
 hypothetical protein rompB [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
 C:Accession: E97835
 R:Ogata, H.; Audic, S.; Renesto-Audiiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
 Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; PMID:21442074; PMID:11557893
 A:Accession: E97835
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1655 <Kur>
 A:Cross-references: GB:AE006914; PIDN:AAL03623.1; PID:gl5620207; GSPDB:GN00173
 C:Genetics:
 A:Gene: rompB

Query Match	11.0%;	Score 331.5;	DB 2;	Length 1655;
Best Local Similarity	24.4%;	Pred. No. 5.6e-09;		
Matches 177; Conservative	81;	Mismatches 240;	Indels 227;	Gaps 40;

QY	13	IQKGLTALFTTTAAIMLTGSG-VLGA-----RT-----VTADGA---ELAAGTNIGRP	59
Db	8	KKLISAGLVTAAGTIVASFAGSAGMAAAIQQRNNTTNAVATTVDGVGFQDTAVPANVAVP	67
QY	60	AGAFVA-----GSTL-----QYTGAF--IVTDADVSVRALDLNNFAAGLFSVTVGDISLGS	107
Db	68	LNAVITAGVNGKLTLPNTAGSNGLEFLNTANNLDVTVRE-----DPTLGF	112
QY	108	VVD-TGGAKLAVNIDDLGLTLTGFG-----TAAYGANPALL-FOGGOAAANNTYTAL	159
Db	113	ITVNVNNAHFNLMLNAGKLTITGOGITNVQAAATKNAVNVQAVNNGAAIDNNDLQGV	172
QY	160	GNITLGGANGL--TIASDPDVLGPITLAGNI-----	189
Db	173	GRIDCGAAASTLVFNLANPTTQKAPLILGDNAVIVNGAGTLNVTNGFIKVSXSKSFATVN	232
QY	190	-----DGGGTTIDNTDA-----AINGTIGNTNPAQAISICAS-----TL	223
Db	233	VINIGDQGGIMF-NTDADNVNTLNLOANGATITFNGTDG-TGRVLVLLSKNAATDFNVTG	290
QY	224	SLGG-----AVIKATTTKLITNAAPVLLTITNANAVLTGAVDNTTG-----DDVGLV	269
Db	291	SLGSLNKGIIEFNTVAVNGQ-----LKANAGANAIVICTNNGAGRAAGFVVSVDNGKVA	344
QY	270	NLNG-----ALSQVNTGNTGNTSLATISVGA-GTATLGGAVIKATTTKLTN-----	314
Db	345	TIDGQVYAKDMVIQSANAVGQVNFRIHVDVTDGTTAFKTAASKVAITQNSNFGTTDFGN	404
QY	315	-ARSVLTLNAVLTGAVDNTTG-----GDNVGV--NLSGALSOVT--CNIGNTNLSLATI	364
Db	405	LAAQIIVPNTMTLNG---NFTGDASPNGTAGVITFDANGTLASADANAVTNNITAI	461
QY	365	NI--GAGVATLDGAVIKATTTKLTDDASVLIFNPVVVTGAIDNT---GNANKGVVIFTG	419
Db	462	EASGAGVQLSGT--HAEELRIGNAGSVFKLADGTGVINGKVNQOTALVGALAAAGTITLDG	519
QY	420	ASTVTDNIGNTAVLAEVSVGAGLLQIGGVKVKANAINLTDNASVVTFTGDSTVTVGSIGGT	479
Db	520	SATITGDIGNA-----GGAAALOGITLANDATKTLTGGANIIGANGG--	562
QY	480	ELPATVNICAGITLRAGSL-----AANNT--DEGA-----ASNLEFNG	516
Db	563	-----TINFOAN-----GGTIKLTSTONNVWDFDLAIAATDQGVVDASSLTTNAQTLTNG	613
QY	517	PACKNKNLTGTIANGNATL--NINAAGTVLA-----NDVSIGTVVAQINTQNKNKIFI	567
Db	614	K-----IGTV--GANKKTLGOFNIGSSKTVILSDGVDAINLELVIGNNGVAVQFAHNTYLT	665
QY	568	NAKNA	572

1
2
3
4

D_b 666 RTNA 670

RESULT 10

AB001110
AD0123
probable autotransporter protein yaph [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0123
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentis,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougl
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; B
Nature 413, 523-527, 2001
A:title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0123
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3705 <KUR>
A:Cross-references: GB:AL90842; PIDN:CAC89847.1; PID:gi15979073; GSPDB:GN00175
C:Genetics:
A:Gene: yaph

Query Match	10.9%	Score 329.5;	DB 2;	Length 3705;
Best Local Similarity	24.8%	Pred. No. 1.7e-08;		
Matches 165; Conservative	88;	Mismatches 233;		
Indels 179; Gaps	33;			

Qy	23	TTSTAAMLTGSG-----VLGAARTVTADGAEAAAGTNGP---GAGAFVAGST	68
Db	1144	TTVTAGELILGANGAFQTSLLDIASGASANTINGYSQTVGAVTNVGTVTLCSGGVLT	1203
Qy	69	LQVTGFTVDADVSVRALDLNFAAGLFSVTGDISLGSVVDTGKANKLAV-----	119
Db	1204	LTNGGILDTGG-----ALMLTAGGASTVAGGUTGAGTTLNINGN-LSVSAANSLSG	1255
Qy	120	--NIDDLGLTTLTGTAAYGANPALLFGQGOAAANNTYTALGNITLGGANAGLTIASDP	177
Db	1256	QTHIADVASVTLTDTGLTQSA-----VEVLGTLNLNGANAAMTNVLSG	1299
Qy	178	D----VLGPTTLIAGNDGGIITDNTDAALINGTIGNTPAAQISIGASTLSLGGAVIKAT	233
Db	1300	DGINTINAAVTLISGNNFSFGAHCQTGDEL--TVGQ---ASNLGASSATVNLG---TLT	1350
Qy	234	TTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALSQV-----TGN	281
Db	1351	SHLILNGVSESIANVLSGVAGSTVDIIGADTALTANNSGFLGQVALAGNSKLTVASTNN	1410
Qy	282	IGNTNSLA-----TISVCAGTATLGGAVIKATTKLTNAAASVLTLT---NAVLTGA	329
Db	1411	LGASSVVALAGAGDTLISLGSFGNTGFSVTSQGVLOVTDDAEV--TLTSSNGSVNAVITDI	1469
Qy	330	VDNTTGGDNNGVVNLSGALSQVGTG---NIGNTNSLATINTG-----AGVATP---	373
Db	1470	ADATLNLDDIALFN-----HVLGTGNLLNVAKNDASTAFDGTGTVGAFSGIVNLNTTF	1524
Qy	374	-----DGAVIKATTKLTDDASVLIFTNPVVVTGAIDNTGNA---NKGVLIETGAS---	421
Db	1525	ALSADNAAALARATLKLSDS-----VTVTGATDRTLHGLDLNGTGLIFDGPSPQSQ	1576
Qy	422	-----TVTONIGNTAVLAEVSVGAG-----LLOIQGGVV---KANAINLTD	459
Db	1577	ANGVTVTDLALNSGTIS--ITGACNWEHEHPVTPPVVLSLEQDQRGDTILLELINAANVTG	1634
Qy	460	NASVVTFTGDSVTGSGIGTELFATVNICAGITIRAGSGSLAANNI--DFGAASNLEFNGPA	518
Db	1635	NANLDDLVDGTAITS--CTQ-----GVESAIOGGSTVANA1HNVGLTSS---NGNG	1682
Qy	519	GK----NYNL--IGTIANGNNATLINAAGTVIANDV-----SIGTVAQINIONKNKIFV	566
Db	1683	GSGLYWNYTLSALELLANGANALLATESG--LTANRVLNABELFGVGGVLV--VDAQNGALT	1740
Qy	567	INAKN	571

Db 1741 ANGN 1745

RESULT 11

AB3528

extracellular serine proteinase (EC 3.4.21.-) [imported] - Brucella melitensis (strain 1

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: AB3528

R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Muler, C.; Los, T.; Ivanova,

.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens

A:Reference number: AB3528; PMID:11756688

A:Accession: AB3528

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2554 <KUR>

A:Cross-references: GB:AE008918; PIDN:AAL53389.1; PID:gl7984282; GSPDB:GN00191

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI10148

A:Map position: II

C:Keywords: hydrolase; serine proteinase

Query Match 10.9%; Score 329; DB 2; Length 2554;

Best Local Similarity 26.2%; Pred. No. 1.2e-08;

Matches 180; Conservative 80; Mismatches 236; Indels 188; Gaps 35;

QY 26 TAAIMLTGSG-VLGAARTVADGAELAGTNI GPAGAFVAGSTLOYTGAFTVTDADV 84

DB 840 TSDIVVTDGLIFNRSDTLNYGGLISGAGFVTSQSGSTTILTCANSYTGATSV----- 892

QY 85 RALDLNNAAGLFSVTGD-----ISLGSVDTGGANKLAVN-IDDGLTL----- 127

DB 893 -----SAGILLVNGDSATGOTS VANGSIL--GGSGIIGGNVVVTDGALAPCSGA 942

QY 128 -TLTGTTAAAGPALLFGGQA-----AANNVTALGNITLGG----- 166

DB 943 GTLTINGSLSAGSILSMQLQAGVAGGALNDLIEVKNLTLGDLTDLVAETAGSYGPG 1002

QY 167 -----ANAGLTASDPDVLGPI-----TLAGN----- 189

DB 1003 IYRLINYTSLDNLGIDGLMPLNGAGAIQTAVAGQVNLLAGTTFNFWGDGVPKFN SAV 1062

QY 190 DGGGIITDNTDAAINCTIGNTPAQISIGASTLSLGGAVIKATTKLTNAAPVLTNA 249

DB 1063 DGGNGTWQSSGNNNWTDATGNINASYSDGAFITGTA---GTVT----- 1105

QY 250 NAVLTGAVDNTTGGDDGVVLNGLALSQ--VTGN-IGNTNSLATISVGAGTATLGGAVIK 306

DB 1106 -----IDNSLG--QVKAEGMFAIDSYAVTGDKLELTGPSTIRVGDGT-TAGAAIY- 1154

QY 307 ATTTLKTNAAVLTLTNA---VLTGAVDNTTGD--NVGVNLSGALSQVGTGNIGNTNS- 360

DB 1155 ATINSVLTNTOLEKTDACTLTLTGA-NSYTGTAINGCTIRISS-----DDNLGVASSD 1208

QY 361 ----LATINIGAVATLDGAVIKATTKLTDDASVLFTNPVVVTCADINTGNAN---KG 413

DB 1209 ISFDGALNTANATDRAIILTGAGTLLTDASTLSLSGPIISGTGALTGKSGTGLLLSG 1268

QY 414 VVIFTGASTVT-----DNIGNTAVLAESVSGAGLQIQGVVVKANAINL 457

DB 1269 TAAHTGGTITAGTLOIGNRGTDSSIDGNLVNNGAL--VFDRAGLTAYTGTI--SGTGTL 1324

QY 458 TDN-ASVVTFTGDSVTGSGIGTELPATVNI GAGITLRAGSL---AANNIDFGAASNLJE 513

DB 1325 TRKNSLTLMWTGTSTVTGE-----TTVSAGT-LALQAGGQIKGTASTLV DGGAEVLID 1376

QY 514 FNG---PAGKNYNLIG-----TIANGNNATLN-----INAAGTVIANDVSTGTV AQINQN 562

DB 1377 GSGSOFATGAGASVVGTVTRDGGTASFDLSLTTSNATGTNTSTIIVA-GSGSOMTQTGT 1435

QY 563 KIFVINAKNADVDIILDAQAISFKGAA 588

DB 1436 ATFGGL-AGTATVDIILDGTMISSGAS 1460

RESULT 12

AF2959

conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AF2959

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AF2959

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1052 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:gl7741659; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu3276

A:Map position: linear chromosome

Query Match 10.7%; Score 323.5; DB 2; Length 1052;

Best Local Similarity 25.0%; Pred. No. 8.5e-09;

Matches 163; Conservative 90; Mismatches 195; Indels 205; Gaps 36;

QY 20 ALFTTSTAI MLTSGVLG-----AARTVTADGAELAAAG-TNIG-----PG----- 59

DB 43 AKLVVEAAATVSGRAIIGRHSAEATVTGDSKWTGDLQVGGTSDPGLAGNTLVN 102

QY 60 -AGAFVAGSTLOYTGAF-----TVTDADVSVRALDLNPF-----AAGLFSVTGDSLGS 107

DB 103 TAGGSV-DSTVAHLGVVAGATGSAIVDGKSVWTVDRNSLEVSGAGSLAVTG-----GG 157

QY 108 VVD-----TGGANKLAVNIDDLTLTLTGTPAAAYCANPALLFOGQAAANTTYTA 158

DB 158 LVDAANIIICTNTGNGSVRVSGADSTVRSRLNVLGYNGSMTVEAGGAVKSRDGYVA 217

QY 159 LGNITLGGANAGLTIASDPDLGPITLAGNIDGGIITDNTDAINGT--IGNTNPAAQI 216

DB 218 ----TVGGSTSAVTVTG-----DSSWAMTGTTFVGYASGAT-- 250

QY 217 SIGASTLSLGGAV--IKATTKLTNAAPVLTLTNANAVLTGAVDNTT---GGDDVGVLNL 271

DB 251 --GNVTVSNGGAI RATGVTGLDLAGASGTMITGAGSKVTAYVDNGTVNSGSDVDFGQG- 307

QY 272 NGALSQVTCN-----IGNTNSLATISVG-----AGTATL---GGA 303

DB 308 SGSLSVVNGGSLDAYNLYVGNALGSSGAVLVSQVSHSVSDGLMVVGNAGNSVEITGGA 367

QY 304 VIKATTKLT-TNAAVLTLT-----NAVLTGAVDNTT---GGDNVGVV----- 342

DB 368 SLAAPTILLIATAGSTGVLSIGAGSGQTARSAGAVEARAIFAGNGSIVFNHSETGYTL 427

QY 343 --NLSGASQVTCNIGNTNSLATINIGAVATLDGAVIKATTK-----LTDDASVLI-- 393

DB 428 SADISGA-CRVVAEAGVT--TLGNNNSYSGTTLISAGMLKGTAKSFGSGGIVNNAELVVDG 485

QY 394 ---FTNPVVVTTGAIDNTGNANKGWIFTGASTVTDNIGNTAVLAESVSGAGLQIQGVV 450

DB 486 GGTLSNAISGTSFEKTGDGN---LLLTGNTSVS---GATA-----VSAGKLVSNGSL- 532

QY 451 KANAINLTNANSVVTTGDSV--TGSIGGTGTELPATVNI GAGITLRAGSGLAANNIDFGA 508

DB 533 -----ASAVSYSGATVGGTGTIG-----GLTVNSGGTGLAPGN-----S 566

QY 509 ASNLEFGPAGKYNLIGTIANGNNATLNINAAGT-----VIANDVSI 551
 | | | | | : | | : | | | : | | | : | | | : | | | :
 Db 567 ICTLTSTGNA-----TFASGSTYAVEIDADGSSDRLAVTGTGTIANDVSL 611

RESULT 13

H98323
hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: H98323
R.;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: H98323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1341 <NR>
A;Cross-references: GB:AE007870; PID:AAK90114.1; PID:gi15160106; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3085
A;Map position: linear chromosome

Query Match	10.7%	Score	323.5;	DB 2;	Length	1341;
Best Local Similarity	25.0%;	Pred. No.	1.1e-08;			
Matches	163;	Conservative	90;	Mismatches	135;	
				Indels	205;	Gaps

Qy	20	ALFTTSTAAIMLGTSGVYL	-----AARTVTADGAELAAG-TNIG-----PG-----	59
Db	332	AKLVEAAAVTSGEALIGRHSASEATVTGDGSKWTTGDLQVGGDTSPDGLAGNCTLNV	-----AAGLSFVSTGDISLGS	107
Qy	60	-AGAFVAGSTLQYTGAF-----	-----TVTDADSVRALDNLNF-----	446
Db	392	TAGGSV-DSTVAHLGVVAGATGSAIVDVGKGSVWTVDRNSLEVSGAGSLAVTG-----	-----GG	446
Qy	108	VVD-----TCGANKLAVNITDDGLTTLTGCTGAAYGANPALLFCGQAAANNITYFA	-----	158
Db	447	LVDAANIIGTNTGGGNSVRVSGADSTVKRSRDLNVGLYNGSMTVEAGAVKSRDGYVA	-----	506
Qy	159	LGNITLGCANAGLTIASDPDVLGPITLAGNIDGGGIITONTDAAINGT--IGNTNPAAGI	-----	216
Db	507	-----TYGGSTSAVTVTG-----	-----DGSSWAMTGTFFVGYASGAT-----	539
Qy	217	SIGASTLSLGGAV--IKATKTTKLTAAPVLTLTNAVALTGAVDNNT-----	-----GGDDVGVNL	271
Db	540	--GNVTVSNGCATRATGVTLGLDLAGASGTWTTIGAGSKVTAVDNCTVNSGVDVGFQG-	-----	596
Qy	272	NGALSOVQTN-----IGNTNSLATSIVG-----	-----AGTATL-----GGA	303
Db	597	SGSLSVYVNGGSLDAYNLYVGNALGSSGAVLSCVGSHSVVDGLMVVYVGNAGNSVEITGGA	-----	656
Qy	304	VIKATTKL-TNAASVLTLT-----	-----NAVLTCGAVDNNT-----GGDNVGVV-----	342
Db	657	SLAAPILINATEAGSTGVLSIGAGSQGTARSAGAEARAIACFAGNGSTVFNHSETGYTL	-----	716
Qy	343	--NLSGALSOVTCNIGNTNSLATINIGAGVATLDGAVIKATTK-----	-----LTDASVLI-----	393
Db	717	SADISGA-GRVVAEAGVT--TLSGNNSYSGCTTISACMLKGTAKSGSGGIVVNAELVVDG	-----	774
Qy	394	---FTNPVVVVTGAIDNTGNANKGVIIFTGASTVTDNIGNTAVLAEYVSGAGLLOQGGVV	-----	450
Db	775	GGTLSNAISGTSGFEXTGDGN--LLLTGNSVYS--GATA-----	-----VSAGKLSVNGSL--	821
Qy	451	KANAINLTDNASVVTTGTGSTV--FGSIGGTFLFATVNICAGITLRAGGSLAANNIDPGA	-----	508
Db	822	-----ASAVSGSGATVGGTGTIG-----	-----GLTVNSGGTGLAPGN-----	855

Oy	509	ASNLEFNPAGKYNLI	GTIANGNATLNINAACT-----	VIANDVSI	551
			: :	:	
		-	- - - - -	- - - - -	
Dd	856	IGTLTSTGNA-----	TFASGSTYAVEIDADGSSDRLAVTGTITIANDVSL	900	

RESULT 14
E97857

Cell surface antigen homolog RC1261 [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97957
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893

A;Accession: E97857
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <KUR>
A;Cross-references: GB:
C;Genetics:
A;Gene: RC1261

Query Match	10.3%	Score 310.5;	DB 2;	Length 652;
Best Local Similarity	24.3%;	Pred. NO. 2.2e-08;		
Matches 166; Conservative	80;	Mismatches 237;	Indels 199;	Gaps 34;

RESULT 15

Query Match	10.2%;	Score	308.5;	DB	2;	Length	256;
Best Local Similarity	31.8%;	Pred.	No. 1e-08;				
Matches	94;	Conservative	43;	Mismatches	104;	Indels	55;
Gaps	13;						
Qy	294	GAGTATGGGAVIKATTTKLTNAAASVLITN---	AVLTGAVDNTTGGDNVGVNLS---	GA	347		
Db							
3	GAGNIT-----	INAAANLNLNNSILFLDCNITTLTGNINNTAGVDSQGQILNLAHDLGS	57				
Qy	348	LSQVTVGNIGNSLATINIGAGVATLGDGAVIKATTKLTDDASVLIFNPVVVTVCAIDNT	407				
Db							
58	SNIIITGDIGNTGSUAAVNVLLGAATLSTILUKATNINLQSNITVLNDDDTVTGNIIDGA	117					
Qy	408	GNANKGVVIFGTGATVDNIGNTAVLAESVSGALLOIQGGVWRKANALNTDNASV--VT	465				
Db							
118	KGVNGN---FTGNAILNGNINN-----	FNILQCGG--NGKILDLQSNITVNSIV	162				
Qy	466	FTGDSVTWTGSGGTGTELFATVNVIGATILR-----	AGGSLAANNIDFGAASLNEFNPGPAGKN	521			
Db							
163	FADSVLAAGTIVSNLS--LDVG-GITFNNSNASSGTLIINTED-----	T	203				
Qy	522	YNLTGTTANGNNATNLINNAAGTVTIANDSVICTVAQINQNNKIFVINKAKNADVDIL	577				
Db							
204	INI--ALLNAIQAKIQIINANIT--NDFSAGDIGDIRIADNTTTTIDAANGNVLL	255					

•
•
•

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:01:34 ; Search time 28 Seconds
(without alignments)
916.923 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: .3022

Sequence: 1 MANISLKLFQAIQKLGKTA.....MIELSLKIIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1510	50.0	2021	1 OMPA_RICCN	Q52657 rickettsia
2	1494.5	49.5	2249	1 OMPA_RICRI	P15921 rickettsia
3	376.5	12.5	1656	1 OMPB_RICJA	O06653 r outer mem
4	375.5	12.4	1654	1 OMPB_RICRI	Q53047 r outer mem
5	333.5	11.0	1025	1 SLAP_CAUCR	P35828 caulobacter
6	333	11.0	1643	1 OMPB_RICPR	Q53020 r outer mem
7	331.5	11.0	1655	1 OMPB_RICCN	Q9kka3 r outer mem
8	299.5	9.9	1645	1 OMPB_RICTY	P96989 r outer mem
9	276.5	9.1	1300	1 L20K_RICRI	P14914 rickettsia
10	259.5	8.6	2003	1 YDBA_ECOLI	P33666 escherichia
11	259	8.6	1953	1 BIGA_SALTY	P25927 salmonella
12	258.5	8.6	928	1 PM10_CHLPN	Q9rb65 chlamydia p
13	251.5	8.3	1286	1 AIDA_ECOLI	Q03155 escherichia
14	250.5	8.3	1039	1 AG43_ECOLI	P39180 escherichia
15	250	8.3	933	1 SLAP_CAMEF	P35827 campylobact
16	249	8.2	1276	1 PMP6_CHLPN	Q92899 chlamydia p
17	243.5	8.1	1569	1 YPUA_ECOLI	P52143 escherichia
18	238.5	7.9	1723	1 PM20_CHLPN	Q92812 chlamydia p
19	236	7.8	1609	1 PM21_CHLPN	Q92605 chlamydia p
20	229	7.6	678	1 YF48_MYCTU	Q10778 mycobacteri
21	228	7.5	1325	1 YDBK_ECOLI	P32051 escherichia
22	224.5	7.4	1672	1 PMPB_CHLMU	Q9pjy2 chlamydia m
23	223	7.4	957	1 Y278_MYCTU	P56877 mycobacteri
24	221	7.3	936	1 PMF7_CHLPN	Q92898 chlamydia p
25	219.5	7.3	333	1 YRP5_FRV6	P18309 chilo iride
26	219	7.2	672	1 PHX5_MOUSE	P08399 mus musculu
27	213	7.0	3591	1 FHAB_BORPE	P12255 bordeletia
28	212	7.0	1288	1 VACA_HELPJ	Q92kw5 helicobacte
29	210	6.9	1150	1 APMU_PIG	P12021 sus scrofa
30	205	6.8	955	1 YCGV_ECOLI	P76017 escherichia
31	205	6.8	1848	1 CBPA_CLOCL	P38058 clostridium
32	204.5	6.8	1754	1 PMPB_CHLTR	O84418 chlamydia t
33	204	6.8	928	1 PMP9_CHLPN	Q9z398 chlamydia p

RESULT 1

ID	OMPA_RICCN	STANDARD:	PRT:	2021 AA
AC	Q52657; P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669;			
AC	Q52670; Q52674;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOMP A) (rOMP A).			
GN	OMPA OR RCI273.			
OS	Rickettsia conorii.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiae; Rickettsia.			
OX	NCBI_TaxID=781;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Malish 7;			
RX	MEDLINE=94171067; PubMed=8125327;			
RA	Crocquet-Valdes P.A., Weiss K., Walker D.H.;			
RT	"Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia conorii (Malish 7 strain).";			
RL	Gene 140:115-119(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Malish 7;			
RX	MEDLINE=21442074; PubMed=11557893;			
RA	Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V., Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., Raoult D.;			
RT	"Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";			
RL	Science 293:2093-2098(2001).			
RN	[3]			
RP	SEQUENCE OF 8-204 FROM N.A.			
RC	STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;			
RX	MEDLINE=97015921; PubMed=8662558;			
RA	Roux V., Fournier P.E., Raoult D.;			
RT	"Differentiation of spotted fever group rickettsiae by sequencing and analysis of restriction fragment length polymorphism of PCR-amplified DNA of the gene encoding the protein rOMP A.";			
RL	J. Clin. Microbiol. 34:2058-2065(1996).			
RN	[4]			
RP	SEQUENCE OF 953-2012 FROM N.A.			
RC	STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;			
RA	Raoult D., Fournier P.E., Roux V.;			
RT	"Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein ompA.";			
RL	Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.			
CC	-!- PTM: GLYCOSYLATED (BY SIMILARITY).			
CC	-!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -			

Q8x8v7 escherichia
P55981 helicobacte
P45508 escherichia
P04949 escherichia
Q48258 helicobacte
P16466 proteus mir
O53553 mycobacteri
P13126 deinococcus
Q48245 helicobacte
O53810 mycobacteri
P42835 saccharomyc
Q48247 helicobacte

ALIGNMENTS

34 203.5 6.7 2660 1 YFEI_ECO57
35 203 6.7 1290 1 VACA_HELPY
36 201 6.7 1250 1 YFAL_ECOLI
37 200 6.6 497 1 FLIC_ECOLI
38 200 6.6 1291 1 VACA_HELPY
39 199.5 6.6 1577 1 HLYA_PROMI
40 199.5 6.6 1501 1 YZ08_MYCTU
41 199 6.6 1036 1 HPI2_DEIRA
42 198.5 6.6 1287 1 VAC2_HELPY
43 198 6.6 801 1 Y747_MYCTU
44 197.5 6.5 1041 1 EGT2_YEAST
45 197.5 6.5 1296 1 VAC1_HELPY

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AB003681; BAA20138.1; -
InterPro: IPR003858; rOmpA_rOmpB.
Pfam: PF02708; rOmpA_rOmpB; 1.
Antigen: S-layer; Cell wall.
CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT CHAIN 528 533 POLY-GLY.
FT DOMAIN 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 12.5%; Score 376.5; DB 1; Length 1656;
Best Local Similarity 24.2%; Pred. No. 1.9e-10;
Matches 184; Conservative 96; Mismatches 269; Indels 211; Gaps 37;

Qy 13 IQKGLKLTFTTAAIMLTGSG-VLGAA-----RT-----VTADGA---ELAAAGTNGPG 59
Db 8 LKLLISAGLVTAATVAVSAGSAMGAAIQONRTTNGVATVVDGQFDQTAAANVAVA 67
Qy 60 AGAFVAGSTLOYTGAFVTADADSVRALDLNFAAGLFSVTGDISLGSVD-TGGANKLA 118
Db 68 PNAVI---TANANGINLTTPAGSFNGFLSNANLAVTVSEDTTLGFINNAANRNF 124
Qy 119 VNIDDLTLTLTGSG-----TAAYGA-NPALLFQGGQAAANNTYALGNITLGGANAGL 171
Db 125 LTLDAKLTTLTGITGIVNVOQAATHNAQIVAKFNGGAAIANNLDSGLGTDFGAAASTL 184
Qy 172 T--TASDDPVLGPTLAGNIDGGIITDNTDAALNGTIGNTPAAQIS-----I 218
Db 185 VFDLANPTQKAPLILADN-----ALIVNGA-----NGTLNVTNGFTQVSKDSFATVKAINI 236
Qy 219 G-----ASTLSLGAIVKATTTKLTNAAPVLTLTNANAVLFG-AVDNTG 262
Db 237 GDGQGFMENTNANALNLAQGTTFNFGTDTGR---LVLLSKNGAATDFNVTGSLG 293
Qy 263 GDDVGLNLGAL--SQVTGNIGNTSLATISVAG-----TATLGG----- 302
Db 294 GNLGIIELNVAINGOLIANAGANAVIGTNGGAAAGPWSVDNGKAATIDGQVYAK 353
Qy 303 AVIKAT-----TKLTNAASVLITLNAVLTGAVD----- 331
Db 354 DMVIOANANGQVFRHIVDVGIDGTTAFKTAASIVAITQNSNFGTDFGNLAAQVTVPD 413
Qy 332 -----NTTG-----GDNVGVNLSG-----ALSVQTNIGNTSLATINI-GAGVATIDG 375
Db 414 TMTLTGNTGDNANPGNTAGVITFAANGCTLASADANAVAVTNITALEASGVGVQLSG 473
Qy 376 AVIKATTTKLDASVLFTNPVVVTGAIDNT-----GNANGVVIPTGASTVTDNIGN-- 429
Db 474 T--HTAELRLGNAGSVFKLADGTVINGKVNQTVLVGGVLAAGAITLDGSATITDIGNGG 531
Qy 430 -TAVLAESV-----GAGLIQGGVVRANA-----INLTD----- 459
Db 532 GGAAQSTLTANDATKTLTGLGAINSIANGTINFPQANGTGIKTSTQNNIVVDCDLAIA 591
Qy 460 -----NASVVTFTGDTVTSIG-----GTELPATNIGAGITLRAAGSGSLAANNIDFGA 508
Db 592 TDQGTGVVDASSLTNAQTLTISGTIGIANGNTTLGQFNIGSKTKTLNGNVAINELVGN 651
Qy 509 ASNLEF-----NGPACKNVLGTIANGNNAIINNAAGTVIANDVSIQVIAQIN 558
Db 652 NGSVQFAHNTYLTTRTTNAAGQRIIFNPVYN-NNTTL---AAGTNLGS---AANPLAEIN 705
Qy 559 IQNNKIFVNAKNADVD-ILDAQAISFKGAASRLFLANVS 597
Db 706 -----FGSKGARADTVLNV-----GEGVNLATNIT 731

RESULT 4
OMP_RICRI

ID AC Q33047; OMPB_RICRI STANDARD; PRT; 1654 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Pollicastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -|- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -|- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -|- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: X16353; CAA34403.1; -
InterPro: IPR003858; rOmpA_rOmpB.
Pfam: PF02708; rOmpA_rOmpB; 1.
Antigen: S-layer; Cell wall.
FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 12.4%; Score 375.5; DB 1; Length 1654;
Best Local Similarity 25.1%; Pred. No. 2.1e-10;
Matches 191; Conservative 98; Mismatches 255; Indels 217; Gaps 41;

Qy 13 IQKGLKLTFTTAAIMLTGSG-VLGAA-----RT-----VTADGA---ELAAAGTNGPG 59
Db 8 LKLLISAGLVTAATVAVSAGSAMGAAIQONRTTNGVATVVDGQFDQTAAANVAVA 67
Qy 60 AGAFVAGSTLOYTGAFVTADADSVRALDLNFAAGLFSVTGDISLGSVD-TGGANKLA 118
Db 68 LNAVI---TANANGINLTTPAGSFNGFLSNANLAVTVSEDTTLGFITVNVNNAHSFN 124
Qy 119 VNIDDLTLTLTGSG-TAAYGA-----NPALLFQGGQAAANNTYALGNITLGGANAGL 171
Db 125 LTLNAGTKLITITGQVTVNAQAATKNAQNVVQVFNNGAAIDNDLKGVRIDFGAPASTL 184
Qy 172 --TIASDDPVLGPTTLAGNIDGGIITDNTDAALNGTIGNTPAAQIS----- 217

```

Db 185 VFNLANPTTKAPLILGDN-----AVIAN-----GVNGTLNVTNGFIQVSNKSFATVKAINI 236
QY 218 -----IGASTLSL--GGAVIKATTTKLNAAPVLT-----LTNANAVLTGAVDN 259
Db 237 ADQOGLIFWTDANNANTLMQAGTTINFTGDTGRLVLLSKHAAATNFN--ITGSL-- 292
QY 260 TTGGDDVGVNLNGAL--SQVTGNICNTNSLATISVGAG-----TATLGGAV 304
Db 293 --GGNLKGVIEFTVAVDGQLTANAGANAVIGTNNAGRAAGFVSVVDNGKVAIDGOV 350,
QY 305 -----IKAT-----TTKLTNRAASVLTLT----- 322
Db 351 YAKDMVIOANATGOVNERHIVDVAGDGTAFKTAASKVTITQDSNFGNTDFGNLAAQIK 410
QY 323 --NAV-LTCAVDNFTG-----GDNVGV--NLSGALSQVT--GNIGTNSLATINI-GAG 369
Db 411 VFNATLIG--NFGDASNPENTAGVITFDANGLESADANAVVNTITAEASGAG 467
QY 370 VATLGGAVIKATTTKLTDASVLIFPNVVTGAIDNT-----GNANKGVVIFGTASTVTD 425
Db 468 VVOLSGT--HAAELRLGNAGSIFKLADGTIVNGKVNOTALVGGALAAAGTITLDGSATING 525
QY 426 NIGN--TAVLAEVS-----VGAG-----LLOIQGVVYKANAIN----- 456
Db 526 DIGNAGGAALQRIITLANDAKKTLTGGANIIGAGGTTIDLQANGGTIKLTQNNIVVD 585
QY 457 -----LTDNASVV--TFTGDSVT-----GSIG-GTELFATVNIAGITTLRAGGSLAAN 502
Db 586 FDLATATDQGVVDASSLNTAQTLLINGKIGTIGANNKTLGQFNKSSGTVLSNGVNAIN 645
QY 503 NIDFGAASNLEF-----NPGAGKNYNLIGTIANGNATNLINAAAGTVIANDVSTIG 552
Db 646 ELVIGNDGAVOFAHDTLYLTRITNAAGQKLIIFNPVN-NGTTL--AAGTNLGS--ATN 699
QY 553 TYAQINIQNKIFV-----INAKNADVDILDAQALSF 584
Db 700 PLAEINFGSKGVNVDVLNAGVEGNLYATNITTTDANVGSF 740

RESULT 5
SLAP_CAUCR
ID SLAP_CAUCR STANDARD; PRT: 1025 AA.
AC P35828; Q46015; Q9RF12;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-layer protein (Paracrystalline surface layer protein).
GN R5AA OR CCI007.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=93007489; PubMed=1393820;
RA Gilchrist A., Fisher J.A., Smit J.K.;
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
RT crescentus paracrystalline surface layer protein.";
RL Can. J. Microbiol. 38:193-202(1992).
RN [2]
RP REVISIONS TO 376; 636 AND 842-843.
RA Awram P.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-JS3001;
RA Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;
RT "The secretion signal of C. crescentus S-layer protein is located in
RT the C-terminal 82 amino acids of the molecule.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Eisen J., Heldberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gehlmann M.L., Haft D.H.,
RA Klonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN [5]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=89008089; PubMed=3049545;
RA Fisher J., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
RT Caulobacter crescentus.";
RL J. Bacteriol. 170:4706-4713(1988).
RN [6]
RP CHARACTERIZATION.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=98292737; PubMed=9620954;
RA Awram P., Smit J.K.;
RT "The Caulobacter crescentus paracrystalline S-layer protein is
RT secreted by an ABC transporter (type I) secretion apparatus.";
RL J. Bacteriol. 180:3062-3069(1998).
CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
CC PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
CC (TYPE I) SECRETION APPARATUS.
CC -!- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF062345; AAC38665.2;
CC EMBL: AF193063; AAF19365.1;
CC EMBL: AE005779; AAK22991.1; ALT_INIT.
CC HSP: P22629; ISWC.
CC TIGR: CCI007;
CC InterPro: IPR001343; Hemlysn-Ca_bind.
CC Pfam: PF00353; hemolysinCabin; 3.
CC PRINTS: PR00313; CABDNCRPT.
CC Cell wall; S-layer; Calcium-binding; Complete proteome.
FT INIT_MET 0
SQ SEQUENCE 1025 AA; 98001 MW; AD7A326E1363D8AC CRC64;

Query Match 11.0%; Score 333.5; DB 1; Length 1025;
Best Local Similarity 24.7%; Pred. No. 1.1e-08;
Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

QY 16 GLKTLFTTAAIMLTGSGVL-----GAARTVTAD-GAELAAGT-----NIGP 58
Db 314 GIETHNVTSGAAILNTSSGVTGLTALNTSGAAGTGTAGAGQNTATTAQAANNVAV 373
QY 59 GAGAFV-----AGSTLQYTGAFVTVDADVSVRALDINNFAAGLFSVTDISL----- 105
Db 374 DGGANVTASTGVTSGTIT--VGANSAAGTGVSVSVANSSTTTTGAITVGTGTAFTAOT 431
QY 106 -GSVD-----TGGANKLVNIDGLLTLTGTCTAAYGANPALLFQGGQAAANN 154
Db 432 AGNAVNTTLTQADVTVTGNSSTTAVTQT--TAAATAGATVAGRVNGAVTITDSAAASAT 489

```

QY 155 TYTALGNITLGGANAGLTIASDPVLPITLAG-----NIDG 191
 Db 490 TAGIATVTLGSGFAA-TI--DSSALTVNLSTGTSLGIRGALTATPTANTILTLNVNG 546
 QY 192 ---GGIITDNTDAA-----INGTIGNTPAAQISIGASTLSLGAVIKATTTKLTNA 240
 Db 547 LTTTGATDSEAAADGGFTTINIAGSTASSTIASLVAADATTLNISGARVTTTSHATAA 606
 QY 241 APVLTFLNA-----NAVLTGAVDNT----- 260
 Db 607 LTGITVTSVGCATLGAELATGLVFTGAGADSIILGATTAIVNGAGDDTVTVSSATLGA 666
 QY 261 ---TGCD--DVGVNLNGALSOVTGIGNTNSLATISVGCATYLG-----G 302
 Db 667 GGSVNGGDDTVLVANVNGSFSADPAFGGFETLRV-----AGAAQOSHANGFTALQLG 722
 QY 303 AVIKATTTKLTNA-----SVLTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVGTIGN 357
 Db 723 ATAGATT--FTNAVNVGLVLAAPTGTTVTLANATGTSDFVNLTLSSNAALAGIVA- 779
 QY 358 TNSLATINIGA-----GVATLDGAVIKATTTKLTDDASVLIFTPVVTGAI DNTGNANG 413
 Db 780 LAGVETVNIATDNTTTHAVDTLTQLTSK-----SIVVTG-----NAG 819
 QY 414 VVIETGASTVDNIGNTAVLAESVSGAGLLOIQGGVVKANAINLTDNASVVTFGDSVT 473
 Db 820 LNL-----TNGNTA-----VTSFASAVTGTGSAVTFYSANTV 854
 QY 474 GSIGGTELFATVNIAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYLNIGTIANGN 532
 Db 855 GEV-----VIRGAGADSLTGSATANDYIIGGAGADTLVYTGCT-----DTFTGGT 901
 QY 533 NATI-NINAAGT-----VIANDVSGTVAQINQNKIFVINAKNADVDILDAQAISPKGA 587
 Db 902 GADIFDINAIGTSTAFVITITDAV-----DKLDLVGISTNGA 939
 QY 588 -ASRLFLANVSL 598
 Db 940 IADGAFGAANVL 951

RESULT 6
 OMPB_RICPR STANDARD; PRT; 1643 AA.
 AC Q3020; Q9ZCM0;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
 DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
 DE OMPB OR SPAP OR SPA OR RP704.
 GN Rickettsia prowazekii.
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP STRAIN=Breil;
 RX MEDLINE=91045972; PubMed=2122457;
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RT "Characterization of the gene encoding the protective paracrystalline-
 surface-layer protein of Rickettsia prowazekii: presence of a
 truncated identical homolog in Rickettsia typhi."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Breil;
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039439; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN=Breil;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNR fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii."
 RL Mol. Immunol. 29:95-105(1992).
 RN [5]
 RP CLEAVAGE SITE.
 RC MEDLINE=92104668; PubMed=1729180;
 RX Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent mutant
 RT deficient in processing."
 RL Infect. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M37647; AAA26390.1; ALT_INIT.
 DR EMBL; AF161079; AAD42234.1; -;
 DR EMBL; AJ235273; CAA15140.1; -;
 DR InterPro; IPR003858; rOmpA_rOmpB.
 DR Pfam; PF02708; rOmpA_rOmpB; 1.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREIL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREIL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREIL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTQEAPLTGA -> INSRSSSVHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 735FDF392E6346CC CRC64;
 Query Match 11.0%; Score 333; DB 1; Length 1643;
 Best Local Similarity 24.4%; Pred. No. 1.9e-08;
 Matches 187; Conservative 85; Mismatches 272; Indels 222; Gaps 38;
 QY 13 IQKGLKLTALFTTTAAITMLGSGV-LGAA-----RTVTAD-----GAELAAAGTNICPGA 60
 Db 8 LKKIISAGLVASTATIVAGFSGVAMGAAMQYNNRTTNAATTFDGGFDQAAGANI-PVA 66
 QY 61 GAFVAGSTLOYTGAFTVTDADSVRLADNFAAGLFSVTGDSLSGVSVDTGANK-LAV 119
 Db 67 PNSVIITANANNPITFNTPNHNSLNSLFDLTANDLA--VTINEDITLGTITTAQQAQKFFNE 124

QY 190 -----DGGIITDNTDA-----AINGTIGNTPAAQISIGAS-----TL 223
 DB 233 VINIGDGGIMF-NTDADNVNTLNLQANGATITENGTDG-TGRVLVLLSKNAATDFENVGT 290
 QY 224 SLGG---AVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGG-----DDGVVL 269
 DB 291 SLGNLGLIIEFTVAVNGO-----LKANAGANAAGVGTNNGAAGFVVSVDNGKVA 344
 QY 270 NLNG-----ALSQVGTGNIGNTSLATISVGA-GTATLGGAVIKATTTKLTN-----314
 DB 345 TIDQOVYAKDWVIOANAVGOVNERHIVDVTGDTGTTAFKTAASKVAITQNSFETTDEN 404
 QY 315 -AASVLTLTNVLGAVDNTTG-----GDNVGVV--NLSGALSQVT--GNIGNTSLATI 364
 DB 405 LAQIIVPNTWTLNG--NFTGDASNPGNTAGVITFDANGTGLASASADANVAVTNNITAI 461
 QY 365 NI-GAGVATLDGAVIKATTTKLTDDASVLTFTNPVVVVTGAIDNT---GNANKGVVIFTG 419
 DB 462 EASGAGVVQLSGT--HAEELKLNAGSVFKIADGTGVNKGNTALVGGALAAAGTITLDG 519
 QY 420 ASTVTDNIGNTAVLAESVSGAGLLQIQGVVYKANAINLTDNASVVTFTGDSVTGSGTGT 479
 DB 520 SATITGDIGNA-----GGAALQGITLANDATKTLTGCAIIGANGG- 562
 QY 480 ELFTVNIAGITURAGGSL-----AANNI--DFGA-----ASNLEFNG 516
 DB 563 ---TINFOAN-----GGTIKLTSTQNNVVDFDLAIATDQTGVVDASSLTNAQTLTNG 613
 QY 517 PAGKNYNLGTTIANGNATL--NINAAGTVIA-----NDVSIGTVAQNIQNNKIFVI 567
 DB 614 K-----IGTV-CANNKTLQGFNIGSKTVLSGDVAINELVIGNNGAVQFAHTYLTIT 665
 QY 568 NAKNA 572
 DB 666 RTTNA 670

RESULT 8

OMPB_RICTY STANDARD; PRT: 1645 AA.
 ID 15-JUL-1998 (Rel. 36, Created)
 AC P96989;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOMP)
 DE (rOMP B) [Contains: 120 kDa surface-exposed protein (Surface protein
 antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
 GN OMPB OR SLP.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wilmingtong;
 RX MEDLINE=94040787; PubMed=8224886;
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
 RT "Cloning and sequence analysis of the gene encoding the crystalline
 surface layer protein of Rickettsia typhi.";
 RL Gene 133:129-133(1993).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN=Wilmingtong;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 the S-layer protein antigens of Rickettsia typhi and Rickettsia
 prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [3]
 RP IDENTIFICATION OF CLEAVAGE SITE.

RX MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 membrane protein of rickettsiae: Identification of an avirulent
 mutant deficient in processing";
 RL Infect. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 DR EMBL: L04661; AAB48987.1; -
 DR InterPro: IPR003858; rOMPb_ompb.
 DR Pfam: PF02708; rOMPb_ompb; 1.
 KW Antigen; S-layer; Transmembrane; Cell wall.
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT TRANSMEM 1354 1645 32 KDA BETA PEPTIDE.
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;
 Query Match 9.9%; Score 299.5; DB 1; Length 1645;
 Best Local Similarity 23.0%; Pred. No. 6.3e-07;
 Matches 174; Conservative 104; Mismatches 275; Indels 205; Gaps 35;
 QY 13 IQKGLKTALETTSTAAMLTSQV-LGA-----ARTVTADGA---ELAAGTNIQGP 59
 DB 8 LKKIISAGLVTAATATVAGSGVAMGVQYNNRTTAAATTVAGDFQDGTGAGVNLPA 67
 QY 60 AGAFVAGSTLQYTGAFVTDADVSVRALDNLNFAAGLSVSTGDISLSQVVD-TGANKLA 118
 DB 68 TNSVI---TANSNNAITFTPNGLNSLFDLTANTLAVTINENTLTGVTNVTQGFNF 124
 QY 119 VNIDDLTLTLTGTAAYGA-----NPALLFGGQAAANNVTYALGNITLGA--- 167
 DB 125 FTGAGKSLTITGHTAQQAATTKSAQNVSVKNAGAAINDNLGSGSIDFTAAPSVL 184
 QY 168 -----NAGLTIASDPDVLGPITLAGNIDGGGI-ITDNTDAAI-----NGTIG 208
 DB 185 EFNLINPTTQEAPLTLGDNKAKVINGANGILNITNGFVKVSKDTFAGIKTINGDNQGLMF 244
 QY 209 NTNP-----AAQISGASTLSLGGAVIKATTTKLTNAAPVLTLTNANAV---LTGAVDNT 261
 DB 245 NTPDPAANALNLQGGNTINFG---RDGTGKL-----VLVSKNGNATEFNVTSGL--- 292
 QY 262 GGDDVGVNLNLN--CALSQVGTGNIGNTSLATISVGA-----TATLGGAV-- 304
 DB 293 GGNLKGVIETFTTAAAGKLIANGAANAVIGTDNGAGRAAGFIVSVDNNGNAATISGGVYA 352
 QY 305 -----IKATTTK--LTNAAS-----VLTLL 321
 DB 353 KDVIQSANAGGOVTFEHLVDVGLGKGTNEKTADSKVIITENASFGSDFGNLAQVIVP 412
 QY 322 TNAVLTGAV--DNTTGGDNVGVNLSGALSQVGT-----NIGNTNSLATINTGAVATLDG 375
 DB 413 NKKILTGNFGDAKNNGTAGVITFNANGTLVSGNTDPNIVVTNKAIEVEGAGIVQLSG 472
 QY 376 AVIKATTTKLTDDASVLTFTNPVVVVTGAIDNTGNANK-----GVVIFTGASTVTDNIGNT 430

Db 473 -- IHGAELRLNAGSIFKLADCTVINGPVNQLVNNALAAAGSILQDGSALITGDIGN 530
 QY 431 AV---LAESV-----GAGLL-----QIOGGVVKANAIN----- 456
 Db 531 AVNALQDITLILANDASKILTLGSGANILIGANAGATHFOANGGTIOITSTQNILVDFDL 590
 QY 457 -LTDNASVY---TFTGDSVT---GSGT-----GTFLFATVNIAGITLRAAGGSLAANNIDF 506
 Db 591 VTTDQTGVVDASSLNNQTLTNGSITGIGANTKTLGRENVSSTILNAGDVAINEEL-- 648
 QY 507 GAASLNFNGPAGKNY--LICTIANGNNATLINNA-----GTVIANDVSTGTVAQINI 559
 Db 649 -----VWENDSGVHLTHNTYITLTKTINAANQGRKIIIVAADPINTDTALADNLGSAE--SP 702
 QY 560 ONNKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
 Db 703 LSNHFAHKAANGD-----SILHIGKGVNLYANNIT 733

RESULT 9
 ID 120K_RICRI STANDARD; PRT: 1300 AA.
 AC P14914;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 120 kDa surface-exposed protein.
 GN P120.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=90136087; Pubmed=2515418;
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
 RT "Cloning, expression and sequence analysis of the gene encoding the
 RT 120 kD surface-exposed protein of Rickettsia rickettsii";
 RL Mol. Microbiol. 3:1579-1586(1989).
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
 CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -!- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
 CC CONFERRING ANTIGENICITY TO THE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X16353; CAA34402.1; --
 CC PIR: S07575; S07575.
 CC InterPro: IPR003858; rOmpA_rOmpB.
 CC Pfam: PF02708; rOmpA_rOmpB; 1.
 CC Antigen: Glycoprotein; Cell wall; S-layer.
 KW FT CARBOHYD 7 7
 FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
 FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
 FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
 FT CARBOHYD 436 436 N-LINKED (POTENTIAL).

FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
 FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
 FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
 FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
 FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
 FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
 FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
 FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
 FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
 FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
 FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
 FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
 FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
 FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
 FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
 FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
 FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
 FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
 SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

Query Match 9.1%; Score 276.5; DB 1; Length 1300;
 Best Local Similarity 25.0%; Pred. No. 5.6e-06;
 Matches 138; Conservative 74; Mismatches 209; Indels 131; Gaps 27;
 QY 98 SVTGISLGSVVDTCGANKLAVNIDDLTLTLTGCTGAAYGANPALLFGGQGAANNVT 157
 Db 7 NATGVNFRHIVDVGA-----DGTAFKT-----AASKVTIT 38
 QY 158 ALGNTLGGANAGLTIASDPDVLGPITLAGNIDGGIITDNTDAAI-----NGTIGNTPA 213
 Db 39 QDSN--FGNTDEG--NLAAQIKVPNAITLTGFTGDSNFGNTAGVITFDANGTLESASAD 95
 QY 214 AQISI-----GASTLSLGGAVIKATTTKLTAAPVLTITNANAVLTGAVDNT-- 261
 Db 96 ANVAVTNNTAIEASGAGVVQLSGT--HAAELRLNAGSIFKLAD--GTVINGKVNOTALV 152
 QY 262 -GGDDVGLNLGALSQVTGNIGNTNSLATI-----SVGAGTATLGGAVI----- 305
 Db 153 GGALAAGTTLDGS--ATIGDIGNAGGAALQRIITLANDAKKTLTGGANILGAGGGTID 211
 QY 306 ---KATTTKLTAASVLTITNANAVLTGAVDNTTGGDNNVNVLSGALSQVT-----GNI 355
 Db 212 LQANGTILKTS-----TQNNIVVD--FDLAIATDQTVVDASSLTNAQTLLTKINGKGTI 264
 QY 356 G-NTNSLATINICAGVATLDGAVIKATTTKLTDDASVLFTNPVVVTGADNTGNANKV 414
 Db 265 GANNKTLGOFNGSSKTVLSNGNVAINELVIGNDGAVQFAHDTYLTIT---RTTNAAGQKG 321
 QY 415 VIF---TGASTVT---DNIGN--TAVLAESVSGAGLQIOGGVVKANAINL-----TDNA 461
 Db 322 IFNPVVNCTTLAAGTNLGSATNPALAEINFGSGKGVNVDVNLVNGECVNLATNITTTDA 381
 QY 462 SVVFTT-----GSTVTVSGIGTTE--LFATVNICAGITLRAAGGSLAAN--NIDFGAASNLEF 514
 Db 382 NVGSFVFNAGGTNIVSGTVGGQGNKFNVALENGTFTVKPLGNATENGNTTIAANSFLQI 441
 QY 515 NGPAGKNYNLIGTIANGNNATLINNAAGTVIANDVSTGTVAQINIQNKKIFVINAKNADV 574
 Db 442 GG-----NY-----TADCVASADGTGIVEFVNTGPIT-----VTLNKQAAPY 478
 QY 575 DILDAQAISFKG 586
 Db 479 NALKQITVSGPG 490

RESULT 10
 YDBA_ECOLI STANDARD; PRT: 2003 AA.
 ID YDBA_ECOLI
 AC P33666; P76087; P76856; P76857; P76859;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

16-Oct-2001 (Rel. 40, Last annotation update)
Hypothetical protein ydbA.
YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA MEDLINE=97251357; PubMed=9097039;
RX Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377 (1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RA MEDLINE=92190338; PubMed=1665988;
RX Moszer I., Glaser P., Danchin A.;
RA "Multiple IS insertion sequences near the replication terminus in
RT Escherichia coli K-12";
RL Biochimie 73:1361-1374 (1991).
CC -/- SIMILARITY: TO S-TYPHIMURIUM ORF NEAR CYSG (AC P25928).
CC -/- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000237; AAC74483.1; ALT_SEQ.
CC EMBL: AE000237; AAC74487.1; ALT_SEQ.
CC EMBL: D90778; BAA15009.1; ALT_SEQ.
CC EMBL: D90778; BAA18880.1; ALT_SEQ.
CC EMBL: D90779; BAA18881.1; ALT_SEQ.
CC EMBL: X62680; -; NOT_ANNOTATED_CDS.
CC EcoGene: EG11307; ydbA.
CC DR Hypothetical protein; Complete proteome.
CC KW CONFLICT 489 I -> V (IN REF. 2).
CC FT CONFLICT 495 I -> V (IN REF. 2).
CC FT CONFLICT 495 I -> V (IN REF. 2).
CC SQ SEQUENCE 2003 AA: 205949 MW: 883A12C8B53220EE CRC64;
Query Match 8.6%; Score 259.5; DB 1; Length 2003;
Best Local Similarity 22.9%; Pred. No. 5.2e-05;
Matches 179; Conservative 93; Mismatches 257; Indels 251; Gaps 44;
QY 19 TALTFTTSTRAIMLTGSGVIGARTVTADGAELAAAGTNIAGAGAGVAGSTLQVYGTFTVT 78
DB 925 TGLVAQSGNSTIINTDSSGII-----DLYGRGSGVGLA---TADTAENOGKITLD 970
QY 79 DADVSVRALDLNNAAGLFSVTGDISLGSVVDTG-----GANKLAVNTDDG- 124

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE OF 1-765 FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=91100301; PubMed=1987123;
 RA Wu J.Y., Siegel L.M., Kredich N.M.;
 RT "High-level expression of *Escherichia coli* NADPH-sulfite reductase:
 RT requirement for a cloned *cysG* plasmid to overcome limiting siroheme
 RT cofactor.";
 RL J. Bacteriol. 173:325-333(1991).
 RC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
 CC in positions 414 and 732.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF133696; AAD39458.1; .
 DR EMBL; AE008859; AAL22340.1; .
 DR EMBL; M64506; AAR27042.1; ALT_FRAME.
 DR EMBL; M64506; AAR27043.1; ALT_FRAME.
 DR PIR; C39200; C39200.
 DR PIR; C39200; D39200.
 DR StyGene; SG10437; bigA.
 KW Virulence; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 1953
 FT
 FT DOMAIN 101 252
 FT REPEAT 101 103
 FT REPEAT 104 113
 FT REPEAT 114 122
 FT REPEAT 123 133
 FT REPEAT 134 144
 FT REPEAT 145 155
 FT REPEAT 156 166
 FT REPEAT 167 177
 FT REPEAT 178 188
 FT REPEAT 189 199
 FT REPEAT 200 210
 FT REPEAT 211 221
 FT REPEAT 222 232
 FT REPEAT 233 243
 FT REPEAT 244 252
 FT CONFLICT 207 207
 FT CONFLICT 514 514
 FT CONFLICT 1698 1698
 FT CONFLICT 1795 1798
 FT CONFLICT 1836 1837
 FT CONFLICT 1836 1837
 SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;
 Query Match 8.6%; Score 259; DB 1; Length 1953;
 Best Local Similarity 22.3%; Pred. No. 5.3e-05;
 Matches 147; Conservative 81; Mismatches 245; Indels 186; Gaps 28;
 QY 32 TCGSVGLGARTTADCAELAGCTNTPGAGAFVA--GSTLYQTGATFTYTDADVSRALDL 89
 DB 394 TCGVISGGDTTWISG-----HSTDVNDATGALISGNGTTFNFGDIASVGGGTAI-IDG 447
 QY 90 NNFA---AGLFSVTGDISLGSVVDTCGANKLAVNTDDGLTLTGTGTAAYGANPALLFQ 146
 DB 448 DNATIKNTGTSISGAGSTGTVID---GNNARNVNDGDMTITDGTGGTGGHITGDNVVVIDNA 504
 QY 147 GCOAAANNNTYTA-----GNITLGGANAGLTIASDPVLGPTTLAGNIDGGGI 194

Db 505 GSTTVSGADATALYTEGDNALVINEGNOTISGAGVGTTRIDGDD-----AHTTNTGDI 556
 QY 195 ITDNTDAA---INGTIGNTPAAQI-----SIGASTLSLG-----GAY 229
 Db 557 AVDGAGSAAVIINGDNGSLTQAGDLLVTDGAMGIITYGTGNEAKNTGNATVRDADSVGEV 616
 QY 230 IKATTTKLTNAAPV-LTLTNAVALTGADVNTTGGDDGVGLNL-----NGALS 276
 Db 617 VAGEKTFKNGKIDVSLNGTCALVSGDMSQVTLDDGINVSVQDSEGVSFSSATGVSVS 676
 QY 277 -----QVTGNTGNITSLATISVAGTATLGGAVIKATTTKLTNAASVLTLTNAVLTGAVD 331
 Db 677 DSNAYDITGNVAISADYGGDDLAAGAPPLTGTVVVGNGNTVT-----LNGALN 724
 QY 332 -----NTTGDNVGVNLSGALSQVGTGNTGNITSLATINICAGVATLDGAVIKATTTKL 385
 Db 735 IDNDLSATGGQYLDVGLS-----VTGDDNDEIDGGINITHSEDPLDG-----TSADI 774
 QY 386 TDDASVLIFTPVVTGAIDNTGNANKGVVIFTGASTVTDN--IGNTAVLAESVSGAGLL 443
 Db 775 T-----GISVSGNS--TVTLNGHSTIDTNTVVGHHVVLARVNNGGSLI 815
 QY 444 -----QIOGGVVKANAINLTD-----NASVVTFTGDSVTVTGSGTGTB----- 480
 Db 816 LGDSDSVVDVNVSVIPTGYTYTNALLMADGEGTSIENKGDITSHGVYSVIRADNGSEVSNS 875
 QY 481 -----LFATVNIG-----AGITLRAGGSLAANNIDFGAASNLFEFGPAGKN-----YNL 524
 Db 876 GDILVIATSSNSEDRAAITRASGEGSAVHNKAGGDTILISDQTPGSGGIEVYPLKWT 935
 QY 525 -----IGTIANGNNATLINAAGTV-----IANDVSI---GTVAQINIQNN 562
 Db 936 HTFYAMMASDYGVDNDEGATIHLOGAGVYGVYATSRGKALNEGNIYLDGLVPTLDDENN 994
 RESULT 12
 PM10.CHLPN
 ID PM10.CHLPN STANDARD; PRT: 928 AA.
 AC Q9RB65; Q9RB64; Q9S6P2; O86163;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Probable outer membrane protein pmp10 precursor (Polymorphic membrane
 DE protein 10) (Outer membrane protein 5).
 GN PM10 OR OMP5 OR CP0303.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=833558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VR1310;
 RX MEDLINE=20007584; PubMed=10539856;
 RA Christiansen G., Boesen T., Hjerno K., Dagaard L., Mygind P.,
 RA Madsen A.S., Knudsen K., Falk E., Birkelund S.;
 RT "Molecular biology of Chlamydia pneumoniae surface proteins and their
 RT role in immunopathogenicity."
 RL Am. Heart J. 138:S491-S495(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;

RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA";
RL Nucleic Acids Res. 28:2311-2314(2000).
[4]
RN SEQUENCE OF 1-914 FROM N.A.
RP STRAIN=CWL029/VR-1310;
RC MEDLINE=99081766; PubMed=9864239;
RX Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RA "Identification of two novel genes encoding 97- to 99-kilodalton outer
RT membrane proteins of Chlamydia pneumoniae";
RT Infect. Immun. 67:375-383(1999).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ133034; CAB37071.1; -
CC EMBL; AE002192; AAF38160.1; -
CC EMBL; AP002546; BRA98657.1; -
CC EMBL; AU001311; CAA04671.1; -
CC PHC1-2DPAGE; O86163; -
CC TIGR; CP0303; -
CC InterPro; IPR003368; Chlamydia_pmp.
CC InterPro; IPR003357; OMP.
CC Pfam; PF02385; OMP; 1.
CC Pfam; PF02415; DUF145; 2.
CC Outer membrane; Signal; Multigene family.
CC SIGNAL 1 25 POTENTIAL.
CC FT CHAIN 26 928 PROBABLE OUTER MEMBRANE PROTEIN PMP10.
CC SEQUENCE 928 AA; 97229 MW; 059D5206A1DD0E1 CRC64;
Query Match 8.6%; Score 258.5; DB 1; Length 928;
Best Local Similarity 23.9%; Pred.No. 2.7e-035;
Matches 147; Conservative 94; Mismatches 235; Indels 139; Gaps 26;
QY 21 LFTTSTAAMLGSGVLGAARTVTADGAELAGNTNIGPGAGFAVGASTLQYTGAFVTDA 80
Db 8 LVLSSTLACFTSCSTVEAA---TAE-----NIGPSPSDGSI NTGTYTPKNTTGI 55
QY 81 DVSVRALDNLNFAAGLFSVTGDISLGSVDVTGGANKLAVNIDGLILTLTGCTGTAAYGAN 140
Db 56 D-----YTLTGDITLNLGDSAAALTGCGFS-DPTESLSFAGKGYSLFN 99
QY 141 PALLFGGQAAANNTYALCNITLGGANAGLTIASDPDLVLPITLGNIDGGGIIT-DNT 199
Db 100 ----IKSSAEGAALSVTDKNLSLTCFSSILFLAARSSVITTPSGKGVKCGGDLTFDN- 154
QY 200 DAAINGTI-----GNTNPAAOISGASTPLS-----GGAVIKATTK 236
Db 155 ----NNTILFKDYCEENGGAISTKNLSLKNNTSGISFECKNSSATCKKGGACATGTV 210
QY 237 LT-NAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALSOVTGNIGNTSLA----T 290
Db 211 ITNNTAPLFSNNIAEAGGAINSTGNCVITGNTSLVFSSENSVTAGNGALSGADVT 270
QY 291 IS-----VGAGTATLGGAVKATTKTKLITNAAISVLTLTNAVLTGAVDNTTGGDNVGVN 343
Db 271 ISGNQVTFESGNOAVANGGAIYAKKLTLASGGGGGSI SNNIVOG---TTAG-NGGALS 325
QY 344 LSGALSOVGTGNIGNTNSLATINIGAVATLDGAVIKATTKTLTDDASV----- 391
Db 326 ILAA-----GEGSLAEGADITFNAGNAVITTPQTTRKNSIDIGSTAKTINLR 373
QY 392 ----LIFTNPVVVTGAIDNTG--NANK-----GVVIFTG-----ASTVTDNI 427

RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
 RA Yamamoto Y., Horiuchi T.,
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ML 308-225;
 RC STRAIN=ML 308-225;
 RA Henderson I.R., Owen P.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 [4]
 RN PRELIMINARY SEQUENCE OF 53-78.
 RC STRAIN=ML 308-225;
 RX MEDLINE=89291704; PubMed=2661530;
 RA Caffrey P., Owen P.;
 RT "Purification and N-terminal sequence of the alpha subunit of antigen
 RT 43, a unique protein complex associated with the outer membrane of
 RT Escherichia coli.";
 RL J. Bacteriol. 171:3634-3640(1989).
 [5]
 RN SEQUENCE OF 53-63.
 RC STRAIN=K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12";
 RL Electrophoresis 18:1259-1313(1997).
 [6]
 RN GENE NAME.
 RP MEDLINE=97257509; PubMed=9103983;
 RX Henderson I.R., Meehan M., Owen P.;
 RA "Antigen 43, a phase-variable bipartite outer membrane protein,
 RT determines colony morphology and autoaggregation in Escherichia coli
 RT K-12";
 RL FEMS Microbiol. Lett. 149:115-120(1997).
 CC -!- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
 CC -!- FUNCTION AS AN ADHESIN.
 CC -!- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
 CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
 CC CHAIN).
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
 CC -!- SIMILARITY: TO ADHESIN AIDI-I AND TO BORDETTELLA PERTACTIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE000291; AAC75061.1; ALT_INIT.
 DR EMBL; D90838; BAA15825.1; ALT_INIT.
 DR EMBL; D90839; BAA15832.1; ALT_INIT.
 DR EMBL; U24429; BAB47869.1;
 DR HSPG; P07505; ISRD
 DR EcoGene; EG12686; flu.
 DR InterPro; IPR004899; Pertact_sup.
 DR Pfam; PF03212; Pertactin; 1.
 KW Outer membrane; Signal; Complete proteome.
 FT SIGNAL 1 52
 FT CHAIN 53 551 ANTIGEN 43 ALPHA CHAIN.
 FT CHAIN 552 1039 ANTIGEN 43 BETA CHAIN.
 FT VARIANT 2 2
 FT VARIANT 41 42 SL -> FF (IN STRAIN ML 308-225).
 FT VARIANT 46 46 SL -> K (IN STRAIN ML 308-225).
 FT VARIANT 157 157 W -> L (IN STRAIN ML 308-225).
 FT VARIANT 188 188 V -> F (IN STRAIN ML 308-225).
 FT VARIANT 303 305 ATN -> STI (IN STRAIN ML 308-225).
 FT VARIANT 320 320 A -> T (IN STRAIN ML 308-225).
 FT VARIANT 372 372 N -> Q (IN STRAIN ML 308-225).
 FT VARIANT 493 493 E -> V (IN STRAIN ML 308-225).
 FT VARIANT 497 497 S -> N (IN STRAIN ML 308-225).

CC	-!	- SIMILARITY: TO ADHESIN AIDI-I AND TO BORDTELLA PERTACTIN.
CC		-----
CC		THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC		or send an email to license@isb-sib.ch).
CC		-----
CC		EMBL: AE000291; AAC75061.1; ALT_INIT.
DR	EMBL:	D90838; BAA15825.1; ALT_INIT.
DR	EMBL:	D90839; BAA15832.1; ALT_INIT.
DR	EMBL:	U24429; AAB47869.1; -
DR	HSP:	P07503; ISRD.
DR	EcGene:	EG12686; flu.
DR	InterPro:	IPR004899; Pertactin_sup.
DR	Pfam:	PF03212; Pertactin; 1.
KW	Outer membrane:	Signal; Complete proteome.
FT	SIGNAL	1 52
FT	CHAIN	53 551 ANTIGEN 43 ALPHA CHAIN.
FT	CHAIN	552 1039 ANTIGEN 43 BETA CHAIN.
FT	VARIANT	2 2 K -> N (IN STRAIN ML 308-225).
FT	VARIANT	41 42 SL -> FF (IN STRAIN ML 308-225).
FT	VARIANT	46 46 T -> K (IN STRAIN ML 308-225).
FT	VARIANT	157 157 W -> L (IN STRAIN ML 308-225).
FT	VARIANT	188 188 V -> F (IN STRAIN ML 308-225).
FT	VARIANT	303 305 ATN -> STI (IN STRAIN ML 308-225).
FT	VARIANT	320 320 A -> T (IN STRAIN ML 308-225).
FT	VARIANT	372 372 N -> Q (IN STRAIN ML 308-225).
FT	VARIANT	493 493 E -> V (IN STRAIN ML 308-225).
FT	VARIANT	497 497 S -> N (IN STRAIN ML 308-225).
FT	VARIANT	

FT	46	46	T	>	K	(IN STRAIN ML 308-225).
FT	157	157	W	>	L	(IN STRAIN ML 308-225).
FT	188	188	V	>	F	(IN STRAIN ML 308-225).
FT	303	305	ATN	>	STI	(IN STRAIN ML 308-225).
FT	308	308	A	>	Q	(IN STRAIN ML 308-225).
FT	320	320	A	>	T	(IN STRAIN ML 308-225).
FT	372	372	N	>	Q	(IN STRAIN ML 308-225).
FT	493	493	E	>	S	(IN STRAIN ML 308-225).
FT	497	497	S	>	N	(IN STRAIN ML 308-225).


```
FT VARIANT 585 585 H -> Y (IN STRAIN ML 308-225).
FT VARIANT 709 709 E -> K (IN STRAIN ML 308-225).
FT VARIANT 721 721 M -> T (IN STRAIN ML 308-225).
FT VARIANT 751 753 GHL -> SHF (IN STRAIN ML 308-225).
FT VARIANT 803 803 S -> P (IN STRAIN ML 308-225).
FT VARIANT 815 815 A -> V (IN STRAIN ML 308-225).
FT VARIANT 824 824 C -> S (IN STRAIN ML 308-225).
FT VARIANT 835 835 LNLVHTS -> MNLINYA (IN STRAIN ML 308-225).
FT VARIANT 845 847 OGT -> LGA (IN STRAIN ML 308-225).
FT VARIANT 855 855 S -> T (IN STRAIN ML 308-225).
FT VARIANT 888 888 Q -> L (IN STRAIN ML 308-225).
FT VARIANT 1025 1025 S -> I (IN STRAIN ML 308-225).
FT CONFLICT 61 63 ETV -> TTT (IN REF. 5).
SQ SEQUENCE 1039 AA; 106841 MW; 5170D647C8DEBE0 CRC64;

Query Match 8.3%; Score 250.5; DB 1; Length 1039;
Best Local Similarity 24.2%; Pred. No. 7e-05;
Matches 150; Conservative 76; Mismatches 230; Indels 163; Gaps 30;

QY 33 GSGVLGAARTVTDGAE-----LAAGTNIG-----PGAGAFVAGSTLQYTGFTV 77
Db 139 GQSLQGRVAVTTLNGGQWHEGAIGATGTVINDKGWQVVRPGT---VATDTVNTGAEGG 195

QY 78 TDADVSVRALDNNFAAGLFSVTCDSLGSVDTGGANKLAVNIDDLGLTLTGTTAA 137
Db 196 PDAB-----NGDTGQF-VRGD-----AVRTINKNGRQIVRAEGT----- 229

QY 138 GANPALLFOGQAAANNYYTALGNTLGG-----ANAGLTIASDPDLGPTLGNIDGG 193
Db 230 -ANTVYVAGDQTVHG--HALDTLNGGYQYVHNGG--TASDIVV-----NSDGWQ 276

QY 194 IITDNTDAAINGTIGNT--NPAAQISIGAS-----TSLGGAVIKATTTKLT----- 238
Db 277 IVKN-----GGVAGNTVTKQGRLOVDAGGTATNVTLKQGGALVSTATVGTGINRLGA 330

QY 239 -----NAAPVLTNLNANA--VLTCADVNTTGGDDVGVNLNGALSQVGTGNIGNTSLATI 291
Db 331 FSVVEGKADNVVLENGRGLDLVLTGHTATNTRVDDGGTLDRVNGGTATTVSMNGGVILLAD 390

QY 292 SVGAGTAT-----LGGAVIKATTTKLTNAASVL-----TLTNAVLTGADVNTTGGDN 338
Db 391 SGAVSGTRSDGKAFSIGGG--QADALMLEKSGSFTLNAGDTATDTTVNGGLFTARGGTL 448

QY 339 GGVVNL-SGALSQVGTGNIGNTSLATNIG-----AGVATLDGAVIKATTTKLTDDASVL 392
Db 449 AGTTTLNNGAILTLGSKTVNNDTL-TIREGDALLQGGSLTNGSVKSGGSLT----- 501

QY 393 IFTNPVVVTGAIDNTGNANKGVVIFTGASTVTDNIGNTAVLAESVVCAGLLQIOGGVKA 452
Db 502 -----VSNITLTQKAVNLNEGTLTNDSTVTDVTDIAQGTGTA--LKLTGSTVLN 547

QY 453 NAI-----NLTDNASVYFTGDSVTGTGSI-----GGTELFATVNT----- 487
Db 548 GAIDPTNVTLASGATWNPONATVQSVVDDLSHAGQIHETSTRTGKVPATLKVKNLNGQ 607

QY 488 GAGTLTRAGGSLAANNIDFGAASNLFPNGPKKNYLNIGTIANGNNAT-LNINAAGVIA 546
Db 608 NGTISLRVPRDPMQANNAD-----RLVIDGGRATGKTTLNLVYNAGNSASGLATSGKIQV 662

QY 547 .NDVSIGTVAQ-INTQNKI 564
Db 663 EAINGATTEGAFVOGNRL 681

RESULT 15
SLAP_CAME
ID SLAP_CAME STANDARD; PRT; 933 AA.
AC P35827;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE S-layer protein (Surface array protein) (SAP).
```

```
GN SAPA.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN=84-32 / 23D;
RA MEDLINE=90354448; PubMed=2387868;
RX Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure.";
RL J. Biol. Chem. 265:14529-14535(1990).
RN [2]
RP ERRATUM.
RA MEDLINE=91035477; PubMed=2229082;
RX Blaser M.J., Gotschlich E.C.;
RL J. Biol. Chem. 265:19372-19372(1990).
CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. THIS PROTEIN IS
CRITICAL FOR VIRULENCE.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
KW EMBL; J05577; AAA23032.1; -.
DR Cell wall; S-layer.
SQ SEQUENCE 933 AA; 96757 MW; F88C729B4BA5B1E9 CRC64;

Query Match 8.3%; Score 250; DB 1; Length 933;
Best Local Similarity 23.2%; Pred. No. 6.6e-05;
Matches 163; Conservative 87; Mismatches 284; Indels 168; Gaps 33;

QY 5 SLKLFQK-----ATQGLKLTALTFTTAAIMLTGSG-VLGAARTVTADGAEL-----AAGT 54
Db 229 TLSAFDKTRWSVLGRDLTNAIFTAITRAALLTQDAELIITKRTNVENINISLDTSGD 288

QY 55 NIGPG-----AGAFVAGSTLQYTGFTVTDADVSVRALDNNFAAGLFSVGTG--DISLGSV 109
Db 289 FVFNQYKGVFNVLGDIVSF-----ATDASKSVNVTGTTTAFATAGTCKVDVVAGKIS 343

QY 110 DTGANKLAVNI---DDGLTILTLTGCTAA---YGANPALLFOGQAAANNVTYALGNI 162
Db 344 ALTADSTRSVNLATNNDTITLTSANAATSVNLKQKQAKDATITTSAMQOKYNNRRNRATI 403

QY 163 TLGCANAGLTIASDPDVLGPTLGNIDGGIITDNTDAAINGTIGNTNPAQISIGAS 221
Db 404 TSATAVENLVKHAATNVA-----LNGG-----MDKLATVTLDNAALTAADIKSAS 449

QY 222 TLSL-----GGAVIKATTTKLTNAAPVLTNLNANA----- 252
Db 450 TLNLINSVNGPKHLYSSKRRYCKFKRAAAKVKLNTTAATDQTVTL---KANATDNSLEF 506

QY 253 --LTGADVNTTGGDDVGVNLNGALSQVGTGNIGNT--NSLATISVG----- 294
Db 507 DSATAKTTSVTASGSGKTLVIKGAETLVNIDTATNALQSVSFGTGGGKFSVKTGT 566

QY 295 -----AGTATLGGAVIKA---TTTKLTNAASVLTLTNAVLTGADVNTTGGDNVGVNL 344
Db 567 GDDKIEFVGTLTLEGSVIDAPGNDTIAMKSAA--LTSANFTMIKNIENVAISDAVATDL 624

QY 345 SGALSO-----VTGNIGNT-----NSLATINIGAGVATLDGAVIKATTTKLTDDASVLI 395
Db 625 SSSAFKNSVIITTKAAADTTLTINKDQVINFTAA---DAGSVKLTIVKLVNDVATLMIV 680

QY 396 NPVVVTGAID-----NTGNANKGVVIFTGASTVTDNIGNTAVLAESVVCAGLLQIOGGV 450
```



```
Db 681 KIVLDAAKDTNIALGTAADKALVIDTGTETL--NI-----TSLVKATSPET 726
QY 451 KANAIN--LTDNASVVTFTGDSVVTGSIG--GTELEATVNICAGITLRAGGSLAANNIDF 506
Db 727 TANTVNAKLTDTVTSII-IDGMQITLGHAGTAGPD-YSKVSMIDASALKAGLTFDASAITL 784
QY 507 GAASNLEFNGPAG-----KNYNLIGTIANGNNATLNINAAGTVIANDVSIGTVAGINIQ 560
Db 785 GA--NATIKGGGADSITVKGNIYVDLYVAGGDDTITLKG-----AEKTDITTVNNFN-A 837
QY 561 NNKIFVINAKN-----ADVDIILDAQAISFKGAAS 589
Db 838 GDKIDIADAKNGTFTFNKITMNSDANLDDYITKAVAGDGSN 879
```

Search completed: April 14, 2003, 16:09:26
Job time : 39 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:07:09 ; Search time 94 Seconds
(without alignments)
1356.842 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKFQKAIQKGLKTA.....MIELSLKIIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.unclassified.*
- 14: sp.vertibrate.*
- 15: sp.virus.*
- 16: sp.bacteriap.*
- 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3022	100.0	619	2 Q9F9H4	Q9F9H4 rickettsia
2	1223.5	40.5	834	2 Q52647	Q52647 rickettsia
3	1103	36.5	1029	2 Q52708	Q52708 rickettsia
4	952	31.5	192	2 Q9F0Q0	Q9F0Q0 rickettsia
5	710.5	23.5	2106	2 Q9XC47	Q9XC47 rickettsia
6	452.5	15.0	2340	16 Q92D91	Q92D91 rickettsia
7	422.5	14.0	1604	2 Q9K999	Q9K999 rickettsia
8	376	12.4	1617	2 Q9KKB9	Q9KKB9 rickettsia
9	372	12.3	210	2 Q9KJH3	Q9KJH3 rickettsia
10	366.5	12.1	1618	2 Q9KKB1	Q9KKB1 rickettsia
11	366	12.1	1654	2 Q93QW9	Q93QW9 rickettsia
12	365	12.1	1616	2 Q9KKB0	Q9KKB0 rickettsia
13	363.5	12.0	1616	2 Q9KKA5	Q9KKA5 rickettsia
14	361.5	12.0	3165	16 Q8XY13	Q8XY13 rickettsia
15	361	11.9	1615	2 Q9KKA8	Q9KKA8 rickettsia
16	360.5	11.9	2691	16 Q8XPUI	Q8XPUI rickettsia

17	359.5	11.9	2751	16 Q8XUK0	Q8XUK0 rickettsia
18	356.5	11.8	2737	16 Q8XPUI	Q8XPUI rickettsia
19	354	11.7	1618	2 Q9KKB4	Q9KKB4 rickettsia
20	354	11.7	3930	16 Q98E20	Q98E20 rickettsia
21	353.5	11.7	667	16 Q8XRM8	Q8XRM8 rickettsia
22	353	11.7	1615	2 Q9F0P9	Q9F0P9 rickettsia
23	338	11.2	3552	16 Q8XSD6	Q8XSD6 rickettsia
24	336	11.1	1643	2 Q9F0P7	Q9F0P7 rickettsia
25	335.5	11.1	204	2 Q9FDP2	Q9FDP2 rickettsia
26	335	11.1	1643	2 Q9F0P6	Q9F0P6 rickettsia
27	331	11.0	3501	16 Q8Y106	Q8Y106 rickettsia
28	330.5	10.9	1616	2 Q9KKB8	Q9KKB8 rickettsia
29	329.5	10.9	197	2 Q56384	Q56384 rickettsia
30	329.5	10.9	2432	12 Q8QZ06	Q8QZ06 chilo iride
31	329.5	10.9	3705	16 Q8ZHA1	Q8ZHA1 yersinia pe
32	329	10.9	196	2 Q9LAQ9	Q9LAQ9 rickettsia
33	329	10.9	2554	16 Q9YDM7	Q9YDM7 brucella me
34	328.5	10.9	204	2 Q9F9R4	Q9F9R4 rickettsia
35	328	10.9	196	2 Q9LAQ8	Q9LAQ8 rickettsia
36	328	10.9	1620	2 Q9KKB5	Q9KKB5 rickettsia
37	327.5	10.8	1616	2 Q9KKA2	Q9KKA2 rickettsia
38	327.5	10.8	1616	2 Q9KKA1	Q9KKA1 rickettsia
39	327	10.8	196	2 Q9LAR2	Q9LAR2 rickettsia
40	326.5	10.8	197	2 Q57548	Q57548 rickettsia
41	326.5	10.8	1616	2 Q9KKB3	Q9KKB3 rickettsia
42	326	10.8	1616	2 Q9KKA0	Q9KKA0 rickettsia
43	325.5	10.8	3705	2 Q9F285	Q9F285 yersinia pe
44	324.5	10.7	197	2 Q52655	Q52655 rickettsia
45	324.5	10.7	204	2 Q9F9R6	Q9F9R6 rickettsia

ALIGNMENTS

RESULT 1

Q9F9H4
ID Q9F9H4 PRELIMINARY: PRT; 619 AA.
AC Q9F9H4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Outer membrane protein A.
GN OMPA.
OS Rickettsia felis (Rickettsia azadi).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales.
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_taxid=42862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21217364; PubMed=11321078;
RA Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,
RA Zavala-Velazquez J.E., Foil L.D., Stothard D.R., Azad A.F.,
RA Walker D.H.;
RT "Rickettsia felis: molecular characterization of a new member of the
RT spotted fever group.";
RL Int. J. Syst. Evol. Microbiol. 51:339-347(2001).
DR EMBL: AF191026; AAG17185.1;
SQ SEQUENCE 619 AA; 60547 MW; 1AA12066EEFA1B CRC64;

Query Match 100.0%; Score 3022; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 8.7e-110;
Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANISLKFQKAIQKGLKTALETTSTAATMLTSGVGLGAARTVTADGAELAGTNIIGPCA 60
DB 1 MANISLKFQKAIQKGLKTALETTSTAATMLTSGVGLGAARTVTADGAELAGTNIIGPCA 60
QY 61 GAFVAGSTLQYTGAFVTVDADVSVALDLNNFAAGLFSVTGDISLGSVYDTGGANKLAVN 120
DB 61 GAFVAGSTLQYTGAFVTVDADVSVALDLNNFAAGLFSVTGDISLGSVYDTGGANKLAVN 120
QY 121 IDDGTLTLTGTTGTAAYGANPALLFGGQAANNNTYALGNITLGGANAGLTIASDPDVL 180

```

RESULT 2
ID Q52647 PRELIMINARY; PRT; 834 AA.
Q52647;
AC Q52647;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Outer membrane protein A (Fragment).
OS OMPA.
GN Rickettsia agari.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=786;
[1]
RN SEQUENCE FROM N.A.
RP RP
RC STRAIN=KAPLAN;
RX MEDLINE=931194085; PubMed=7680636;
RA Gilmore R.D.Jr.;
RT "Comparison of the ompA gene repeat regions of Rickettsiae reveals
RT species-specific arrangements of individual repeating units.";
RL Gene 125:97-102(1993).
DR EMBL; L01461; AAA26387.1; -.
FT NON_TER 1
FT NON_TER 834
SQ SEQUENCE 834 AA; 79350 MW; 703F3BAE9B51176C CRC64;

Query Match 40.5%; Score 1223.5; DB 2; Length 834;
Best Local Similarity 51.3%; Pred No. 3.5e-40;
Matches 307; Conservative 46; Mismatches 93; Indels 153; Gaps

QY 20 ALFTTSTAAMITGS-----GVL---GAARTVTAD--GAELAAGTNIGPC----A 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 SVLTITNANAVLTGAIDNTTGVNDVGNVLNCGALSQVGTNGTNTSLATISVGACTATLG 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 GAFVAGSTLQYGA---FTVTADV-----SVRALDNNFAAGLFSVGTGIS 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 GAVIKRTTDLTNAASVLTLTNNANAVLTGAIDNTTGVNDVGNVLN---GALSQVGTNI- 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT: 3
Q52708
ID Q52708 PRELIMINARY; PRT; 1029 AA..
AC Q52708;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Outer membrane protein A (Fragment).
OS OMPA.
NCBI_taxid=781;
GN Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
OX NCBI_taxid=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENYA TICK TYPHUS;
RX MEDLINE=93194085; PubMed=7680636;
RA Gilmore R.D. Jr.;
RT "Comparison of the ompA gene repeat regions of Rickettsiae reveals
RT species-specific arrangements of individual repeating units.";
RL Gene 125:97-102(1993).
DR EMBL; L01462; AAA99908.1; -.
FT NON_TER 1
FT NON_TER 1029
SQ SEQUENCE 1029 AA; 99999 MW; 09573881A5B9BDC1 CRC64;

Query Match 36.5%; Score 1103; DB 2; Length 1029;
Best Local Similarity 42.3%; Pred. No. 1, 9e-35;
Matches 283; Conservative 60; Mismatches 194; Indels 132; Gaps

Qy 16 GLKTLFTTSTAAMIT- GSGVLGAARTVTADGAELAGTNGIGAGAF-----VAGSTL 69
| | | | | | | | | | | | | | | |
Db 365 GNTNALATISVGAGKATGLGAIKATTTKLTDNASQVTFNPVVVTGAIDTGNNGIV 424

Qy 70 QYTGATVTDADVSVALDINNPAAGLFSVTGD-----ISLGS 107
: | | | | | | | | | | | | | | |
Db 425 TFTGGSTVGTGNTGNALATVNVGAGLLRVGGVVKSNITLTDNASQVTFNPVVVTGA 484

Qy 108 VDTGGANKLAVNIDGTLT-----LTGTGTAAYG----- 138

```

```
Db 485 IDNTGNANNGIVTFTGDSVTGNTGNATATISVGAGKATLGGAIKATTTKLTNDASQ 544
QY 139 ---ANPALL-----FQGOAANN--TYTALGNITLGGANAGL----- 171
Db 545 VTFTNPVVVGTGADNTGNANNGIVTFTGDSVTGNTGNATATVNVG---AGLLRVGG 601
QY 172 -----TI-----ASDPDLVLPITLAGNID-----GGGIITDNTDAAINGTINPAAQI 216
Db 602 VYKSNITNLTNDASQVTFNPPVVVGTGADNTGNANNGIVTFTGDSVTGNTGNATATV 661
QY 217 STAGSLGGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALS 276
Db 662 NVGAGLLRVGGVKSNTNLTNDASQVTFNPP--VVVTGAIDN--TGNANNGIVTFTGN--S 718
QY 277 QYTGNTGNATATISVGAGTATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTT 334
Db 719 TVTGNIGNTALATVNVGAGTATLGGAVIKATTTKLTNAASVLTLTNNAVLTGAIDNT 778
QY 335 GGDNVGVNLGALSQVGTGNTGNATATINICAGVATLDGAVIKATTTKLTDDASVLIF 394
Db 779 GVDNVGVNLGALSQVGTGNTGNATATISVGAGKATLGGAVIKATTTKLTNDASQVTF 838
QY 395 TNPVVVGTGADNTGNANNGIVTFTGASTVDNIGNTAVLAESVSGAGLIQIOGVVYKANA 454
Db 839 TNPVVVGTGADNTGNANNGIVTFTGASTVDNIGNTALATVNVGAGLLRVGGVYKSN 898
QY 455 INLTNDAS-----VVTFGTGDSVTGSGTGTATVATVNIAGI 491
Db 899 INLTNDASAVTFTNPVVVGTGADNTGNANNGIVTFTGDSVTGNTGNATATISVGAGK 958
QY 492 TLRAGSGLAANNIDF--GAASNLEFNGPAGKNYNLGNTIANGNNATININAAGTVIANDVS 550
Db 959 ATLGGAIKATTTKLTNDASQVTFNPPVVV--CAIDNTGNANNGIVTFTGDSVTGNTGN 1017
QY 551 IGTVAQINI 559
Db 1018 TNALATVNV 1026

RESULT 4
Q9F0Q0 PRELIMINARY; PRT; 192 AA.
AC Q9F0Q0;
DT 01-MAR-2001 (TREMREL.16, Created)
DT 01-MAR-2001 (TREMREL.16, Last sequence update)
DE Outer membrane protein OmpA (Fragment).
OS OMPA.
GN Rickettsia sp. California 2.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=147259;
RN [1]
RP SEQUENCE FROM N.A.
RA Roult D.;
RC STRAIN=CALIFORNIA 2;
RT "A new SFG rickettsia isolated from fleas.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CALIFORNIA 2;
RA Roult V., Roult D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF210694; AAG48555.1;
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 18596 MW; A9E1C76C41D4590D CRC64;

Query Match 31.5%; Score 952; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 2.6e-30;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LFQKAIQKGLKLTFTTTAAIMLTGSGVLGAARTVTADGAELAAAGTNPICGAGAFVAGS 67
```

```
Db 1 LFQKAIQKGLKLTFTTTAAIMLTGSGVLGAARTVTADGAELAAAGTNPICGAGAFVAGS 60
QY 68 TLQYTGAFVTVDADSVSRALDNLNFAAGLFSVTGDSISLGSVVDTCGANKLAVNIDGLTL 127
Db 61 TLQYTGAFVTVDADSVSRALDNLNFAAGLFSVTGDSISLGSVVDTCGANKLAVNIDGLTL 120
QY 128 TLTGTTGTAAYGANPALLFQGGQAAANNVTYALGNITLGGANAGLTITASDPDVLGPITLAG 187
Db 121 TLTGTTGTAAYGANPALLFQGGQAAANNVTYALGNITLGGANAGLTITASDPDVLGPITLAG 180
QY 188 NIDGGGIITDNT 199
Db 181 NIDGGGIITDNT 192

RESULT 5
Q9XC47 PRELIMINARY; PRT; 2106 AA.
AC Q9XC47;
DT 01-NOV-1999 (TREMREL.12, Created)
DT 01-MAR-2001 (TREMREL.16, Last sequence update)
DT 01-DEC-2001 (TREMREL.19, Last annotation update)
DE Outer membrane protein A.
GN OMPA.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RA Stenos J., Walker D.H.;
RC STRAIN=PHS;
RC MEDLINE=20487299; PubMed=11034486;
RT "The rickettsial outer-membrane protein A and B genes of Rickettsia australis, the most divergent rickettsia of the spotted fever group.";
RL Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
DR EMBL; AF149108; AAD39531.2;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
SQ SEQUENCE 2106 AA; 209075 MW; A12B28F54BEEDECC CRC64;

Query Match 23.5%; Score 710.5; DB 2; Length 2106;
Best Local Similarity 25.5%; Pred. No. 5.6e-20;
Matches 300; Conservative 79; Mismatches 217; Indels 579; Gaps 42;

QY 1 MANISLKLFQKAIQKGLKLTFTTTAAIMLTGSGVLGA-----RTVTADG 47
Db 1 MANISPKLFQKAIHKSLKALFTTTAAIMLSSSGAWGAAGVISVNDAAFSNRVANW 60
QY 48 AELAACTNTPGAGAFVAG-----STLQYTGAFVTVDADSVSRALDNLNFAAG---LFSVT 100
Db 61 NEITAG---GAANGNHADGPQDNEAFYGGNHTIT--ADEAGRIITAINVAGTTPVALNST 116
QY 101 GDISLGSVVDTCGANKLAVNIDGLTLTLTGTTAAAYGANPALLFQGGQAAANNVTYALG 160
Db 117 QNTSVGSIV-TGG-NLLPVTIADGKSLTLTGTKAVALD-----HGFDAAADN--YTLGL 166
QY 161 NITLGGANAGLTITASDPDVLGPITLAGNIDGG-----GIITDNTDAINGTIGNT----- 210
Db 167 AIALGGATAGLTQSATPAL--INLAGAIDGNNNDHGEITVNTRTSFTGIVGKTTHAV 224
QY 211 -----NPAQISIG----- 219
Db 225 TFNNGNAGGVSTAGAISATVVMIGEDAGNVASVVQTLTGAVNFADGALTANNIAGA 284
QY 220 -----ASTLSLG-----AVIKATTTKLTNAAPVLT 246
Db 285 VTANNTTGLTVGAGDVTGAICVNGGNILKQVLFNGASNVATIDATNTYINNAANVTA 344
QY 247 T-----NANAVLTGAV---DNTTG-----GDDVGVNLNGA--LSQV-- 278
Db 345 AGAITAANFAADGALTANNIAGAVTTANNNTGTLTVGAGDVTGAICVNGGNILKQVLF 404
```

```

QY 279 -----TCNIG-----NTNSLAT 290
Db 405 NGASNVATIDATNVTINNAANVTAGAITAAVFAADGALTANNGIAGAVTTANNCTG 464
QY 291 ISVGAGTAT-----LGGAVIK-----ATTIKLTNAASVLTLTNAV----- 325
Db 465 LTVGAGDVTGALGVNGNLLKQVLFNGASNVATIDATTVTINNVAANVTAGAITAAVNF 524
QY 326 -----ITGAV-----DNTTGGDNVGVNLSGALSQVTGNI----- 355
Db 525 AADGALTANNGIAGAVTTANNNTGTLTVGAGDVTGAGVNGNLLKQVLFNGASNVATID 584
QY 356 -----GNTNSLATINIGAVAT----- 372
Db 585 ATNVTINNAANVTAGAITAAVFAADGALTANNGIAGAVTTANNNTGTLTVGAGDVTG 644
QY 373 ---LDG-----AVIKATTKLTDDAS-----VLFTN 396
Db 645 AIGVNGNLLKQVLFNGASNVATIDATTVTINNVAANVTAGAITAAVFAADGALTANN 704
QY 397 PVV-----VTGAID-NTGN----- 409
Db 705 GIVGAVTTANNNTGTLTVGAGDVTGAGVNGNLLKQVLFNGASNVATIDATNVTINNA 764
QY 410 -----ANKGV----- 414
Db 765 ANVTAGAITAAVFAADGALTANNGIAGAVTTANNNTGTLTVGAGDVTGAGVNGNLL 824
QY 415 ---VIPTGASTV-----TDNIG----- 428
Db 825 KQVLFNGASNVATIDATNVTINNAANVTAGAITAAVFAADGALTANNGIAGAVTTAN 884
QY 429 -NTAVLAESVGA-----GLLOIOGG-----VVKANAINLTDNASVVTFTG 468
Db 885 NNTGTL-----TVGAGDVTGAGVNGNLLKQVLFNGASNVATIDATTVTINNVAANVTAG 941
QY 469 DST-----VTGSIG----- 477
Db 942 AITAAVFAADGALTANNGITGAVTAAGGNGTLTAGVGTGAVGTNNAASKVLNAKVDN 1001
QY 478 -----GTELEPA-TVNI-----GAGITLRAGGSAAANNIDFGA-----ASNLEFNGPAGK 520
Db 1002 GAADLVFTSDIYARTVNFDFDTAGGAGVGTIOVGGNLIATNVNFGGAGGGLXELNGPVGK 1061
QY 521 NYNLIGTIANGNATLAINAAGTVIANDVSTGTVQAOINQONKLFVINAKNADVDILDAQ 580
Db 1062 SYTLSTGIANGNATINILGLTVBNAIXIGTVQAOINQONKTFAINVKNAIDIEILNAQ 1121
QY 581 AISPKGAASRLFLANVLSQMIELSLKLIYPVILT 615
Db 1122 AIDFKGANSKFLVNSATDTRVVTIKNDLPAPAT 1156

```

RESULT 6

```

Q9ZD91 ID Q9ZD91 PRELIMINARY; PRT; 2340 AA.
AC Q9ZD91;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Cell surface antigen (SCA3).
GN RP451.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

```

```

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "the genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria";
RL Nature 396:133-140(1998).
DR EMBL; AJ235271; CAA14908.1;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Complete proteome.
SQ SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;

```

```

Query Match 15.0%; Score 452.5; DB 16; Length 2340;
Best Local Similarity 28.5%; Pred. No. 5.7e-10;
Matches 194; Conservative 84; Mismatches 232; Indels 171; Gaps 34;

```

```

...QY 26 TAAIMLTGSGVLGAARTVTADGAELAAAGTNGPGAGAFVAGSTLOYTCGAFVTD---ADV 82
Db 407 TDCVTFNNNSVNGGTLIIINAKNTISAKLLNATKAIQINANLTMNHPGAGDISIRIADN 466
QY 83 SVRALDLNFAAGLFSVTGDISL---GSVDTGGANKLAVNIDDLGLTLTLTGTTGTAAYGA 139
Db 467 TIYTIADKN-----GNVLLNNAKIIFEGADSMALINTGVTADRTFT----- 510
QY 140 NPALLFOGQQAANNTYTALCNITLGGANAGLTIASDPVLGPITLAGN-----ID 190
Db 511 ----IYNLNOSGNDY---GIVKIEAIKVVITANOS---GPYITIGDNTHRLKELIVE 560
QY 191 GGG-IITDNTDAAINGTIGNTPAAQISIGASTLSLGA---VIKATTTKLTNAAPVLT 246
Db 561 GAGDIIDDT---IFTKLLSINSTGQITFN-RTLDGAGGNIAGKKGHTLVNG----- 610
QY 247 TNANAVLTGAVDNTGGDDVGLNMGALSQVTCNIG-NTNSLATISVAGTATLGGAVI 305
Db 611 -----VTGST--TSENNOGILTINS--GNITGVIGTNEGLKLVNIGADPVTCANVF 660
QY 306 KATTTKLTNAASVLTLTNAV-LTGAVDNTGGDNVGVNLSGALSQVTCNIG-NTNSLAT 363
Db 661 --ASVALTNPSVLLIADGVTLTGEV--TTHNNTKGVLSL-GTGSNITGQITGTSAALEK 715
QY 364 INIGAVATLDGAVIKATTTKLTDDASVLIPTNPVVVTGAIDNTGNANKGVVIFTGASTV 423
Db 716 INIGAGASNID-SNIYAGSTVLTDTSELTLNNDVVVNSNIITTAGNSCKLIFTNGGI 774
QY 424 TDNIG-NTAVLAEV-----SVGA-----GLLOI 445
Db 775 TGNIGANGAALQEVFVNGTNTNIGGTANSQNTVAHSAANVVTGLTTGALKYKDTGTIIA 834
QY 446 QGVV-----KANAINLTDNASV-----VFTFGDSTVTGSIG--GTELEFA 483
Db 835 HGLVGDIDENKAGKILGDGAMIDGSVLCNGGVAGTLDFIGDGNVTQNIAGDNANSIS 894
QY 484 TVNIGA-----GITLRAGGSAAANNIDFGA--ASNLEFNGPAGKN 521
Db 895 TINIQDNTKNVTIANDIFVDNIHFTNGGTLQGLNLTTHNIDFGANGGTLEFNG--NNT 952
QY 522 YNLIGTIANGNATLAINAAGTVIANDVSTGTVQAOIN-----QNNKIFVINAKNADVD 575
Db 953 YNLAIIVNCQNGIL--NAFTNLKASDDTGTIVKLIINIGOIGTPQN---FTIQVNNKLT 1007
QY 576 ILDA--QAI5FKGAASRLFLA 594
Db 1008 LVSSVNSSINFGDANSQILLS 1028

```

RESULT 7

```

Q9KK99 ID Q9KK99 PRELIMINARY; PRT; 1604 AA.
AC Q9KK99;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE OmpB (Fragment).
GN OMPB.
OS Rickettsia helvetica.

```

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_TaxID=35789;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=C9P9;

RA MEDLINE=20393643; PubMed=10939649;

RA Roux V., Raoult D.;

RT "Phylogenetic analysis of members of the genus Rickettsia using the
 gene coding the outer-membrane protein ompB (ompB).";

RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).

DR EMBL: AF123725; AAF34128.1; -

DR InterPro: IPR003858; rOmpA_rOmpB.

DR Pfam: PF02708; rOmpA_rOmpB; 1.

FT NON_TER 1604 1604

SQ SEQUENCE 1604 AA; 162714 MW; 83F9AE7597837B32 CRC64;

Query Match 14.0%; Score 422.5; DB 2; Length 1604;
 Best Local Similarity 29.7%; Pred. No. 5.7e-09;
 Matches 194; Conservative 83; Mismatches 230; Indels 147; Gaps 38;

QY 37 LGAARTVTADGAELAACTNIG-PCAGAFVAGSTLOYT---GAFTVT-----DADSV 84

Db 1 MGAA-----NRDINANVLDVIGICVITVGGSLTYVVGAGGCTATATGNDNGNATV 54

QY 85 RALDLNFAAGLSVTG-DISLGSVVDTGANKLAVNIDDLTLITGT-GTAAYGAN-- 140

Db 55 LFGSINLQNGVFAVAGADIAIGSVGTAG-QLLTVNI-AGNTLTNGAPVAAFPNTY 112

QY 141 -----PALLFOGQQA--ANNYTYAL-GN-TLGGANAGLTITASDPVLGPITLA 186

Db 113 TNLGPNVFDAAAFKVSLSAAGADGKTATFNGTATFNGTAAGKII----- 160

QY 187 GNIDGGIITDNTDAATNGTIGNTPA-----AQISIGASTL---SLGGAVIKATTT 235

Db 161 -NIDNGNI-----AFNGTIGNNGIOGLTGTGNAQATLNANKFDTGVCAGSVSLDNG 213

QY 236 KLTNAAPVLTLNANAVLTGAVDNTTGGDDVGLNLNGALSQVGTGNTGNTSLATISVGA 295

Db 214 SILNVADGVNITGAVTNISIDGKAAND-GTVNFGLD-SAVSTDIGANNAATAVNV-A 269

QY 296 GTATL-----GGAVIKATTKLTNAASVLTIL--NAVLTGAVDNTTGGDVGVLNLSGA 347

Db 270 GELTFQGGNNGGGTTFQAQINLTAGSLIKFTQDHAVTGNLTNTSGTNNOGGFSVNG 329

QY 348 LSQVGTGIG-NTNSLATINIGAGVATLDGAVIKATTKLTDDASVLFTNPVVVTGAI-- 404

Db 330 DVFVTGIGAGNSLATINFE-----TDD-SLIVHKAATNATNALFI 370

QY 405 --DNTGNANKGVVIFTGAS-TVTDNIG-NTAV-----LAESVGA--GLLIQIOGGVVYKANA 454

Db 371 QNYNTATANTGILKLGHTGYAINGNIGANNALKLVDLADDNGAATNFTLKQSSSIKAQN 430

QY 455 INLTD-NASVVTFTGDSVTGSI-----GGPELF--ATVNIGACITLRAGGSL--AA 501

Db 431 ISLADQDNLTLEEGTTITGDIINTKNGNGTIVLTGNATNGGTGVDIAAGGNLGRIT 490

QY 502 NNIDFGAASNLFPNGPAGK-----NYNLIGTIANG-----NNATLINAACTVIANDVSI 553

Db 491 KNLNLG-KDLSYGGTGVKVSAPNGINFAANETLSLNTTDPILLASDIITTDKQ-GI 548

QY 554 VAQINTONKIFVINAKNDVILDQAQATSFKAASRLFLANVLSQMLIELSLK 607

Db 549 IDGSALTNDQTLTIAGKIGVDIINNPA-----QNVALQOLKIGSSK 590

RESULT 8

O9KKB9

ID O9KKB9 PRELIMINARY; PRT: 1617 AA.

AC O9KKB9;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE OmpB (Fragment).
 GN OMPB.
 OS Rickettsia aeschlimannii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_TaxID=45262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC16;
 RX MEDLINE=20393643; PubMed=10939649;
 RA Roux V., Raoult D.;

RT "Phylogenetic analysis of members of the genus Rickettsia using the
 gene coding the outer-membrane protein rOmpB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 DR EMBL: AF123705; AAF34108.1; -
 DR InterPro: IPR003858; rOmpA_rOmpB.
 DR Pfam: PF02708; rOmpA_rOmpB; 1.
 FT NON_TER 1617 1617
 SQ SEQUENCE 1617 AA; 163501 MW; B7A1B5DB1403BE62 CRC64;

Query Match 12.4%; Score 376; DB 2; Length 1617;
 Best Local Similarity 25.0%; Pred. No. 3.6e-07;
 Matches 168; Conservative 80; Mismatches 241; Indels 184; Gaps 33;

QY 38 GAARTVTADGA-----ELAACTNIGPGAGAFVAGSTLOYTGAFVTVDADSVRALDL---NN 91

Db 13 GAATTV-DGAGFDQTAAPANVAVALNAVI---TANANGINLNTPDGSEFNGFLDTANN 67

QY 92 FAAGLSVTDISLGSVVD-TGGANKLAVNIDDLTLITGTG-TAAYGA-----NPAL 143

Db 68 LAV---TVSADTTLGFTITNAANNANSFDLRVDAGKLTITGGQITTAQAAVTKNAQNVVA 124

QY 144 LFOGQQAANNNTYALGNITLGGANAGL--TIASDDPVLGPTPLAGNIDGGGIITDNTDA 201

Db 125 QFNGGAAIANNDLSGVTIDLGAASATLVFNLANPTTKAPLVLGN---AVIAN--- 176

QY 202 AINGTIGTNPAAQIS-----IG-----ASTLSLGGAVIKATTT 235

Db 177 GYNGTLNVINGFIOVSDKSFATVKTINIGDQGFIENTDATVGNALNLQAGGATINFGT 236

QY 236 KLTNAAPVLTLNANAVLTGAVDNTTGGDDVGLNLN--GALSQVGTGNTGNTSLATISV 293

Db 237 DGTGR--LVLLSKHGAANDFNVTGSLGGLKGVIAFNTVAVAGQLIANAGPANAVIGTN 294

QY 294 GAG-----TATLGG-----AVIKAT-----TTKLTNAA 316

Db 295 GAGRAAGFVVSDNGNAATIAQOVAKOMVIOANAGGEVNFHIVDVGTDGTAFKTA 354

QY 317 SVLTLTNAVLTGAVD-----NTTG-----GDNVGVVNLSG-----ALS 349

Db 355 SKVAITONSFGTDFGNLAAQITVPTWTLTGNTFTGDSNHNAGTAVITFAANGTLASA 414

QY 350 QVTGNTGNTSLATINI-GAGVATLDGAVIKATTKLTDDASVLIFPTNPVVVTGALDNT- 407

Db 415 SADANVAVTNNITAEASGIVGVQLSGT--HTAELRLGNAGSIFKLADGTVINGKVNQTA 472

QY 408 ---GNANKGVVFTGASTVTDNIGTAVLAESVSGAGLLQIOGGVVKANILNTDNASV 464

Db 473 VYGGALAAAGATLDGATITGIGN-----GGVAAALOG-----I 507

QY 465 TFTGDSVTGSIIGGTGELFATVNIAGITLRA--GGSLAANN-----IDFGAASNLFPNGP 517

Db 508 TLANDATKTLTGGANILIG-ANVGGTIDQANGGKIKLTNTONNIIIVDFDLAIATDQTV 566

QY 518 A-----GKNYNLTGTTA--NGNATL---NINAA-----GTVIANDSVSTGVAQINI 559

Db 567 VDASSLTNAQTTLTISGTGAIGANNKTLQGFNIGSKTALNGVAINELVICNGSVQF 626

QY 560 QNNKIFVINAKNA 572

Db 627 AHNTYLTITSTNA 639

RESULT 9

Q9KJH3 PRELIMINARY; PRT; 210 AA.
 AC Q9KJH3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE Outer membrane protein A (Fragment).
 OS Rickettsia sp. FUJ98.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=116094;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FUJ98;
 RA Zhang P.H., Cao W.C., Zhang X.T., Xu R.M., Dai X.H., Gao D.Q.;
 RT "Identification of Spotted Fever Group Rickettsiae in Ticks from
 Southern China".
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF169629; AAF86358.1;
 FT NON_TER 210
 SQ SEQUENCE 210 AA; 5BA5DA67DAACC82B CRC64;

Query Match 12.38; Score 372; DB 2; Length 210;

Best Local Similarity 50.28; Pred. No. 7.1e-08;
 Matches 113; Conservative 15; Mismatches 63; Indels 34; Gaps 11;

Qy 1 MANISLKLFOKAIOGLKLTFTSTAAIMLTGSGVLGAARTVTA-----DCAE--- 49
 Db 1 MANISPKLFOKAIOGLKLTFTSTAAIMLTSSGVLGAAGVATNDAAAFSNDAAEANN 60
 Qy 50 ----LAAGTNGPGAGAFVAGSTLTQYGAFTVDADSVRALDNNFA----AGLFSVTG 101
 Db 61 WDELTAEGVANGIPAGGQNNWAFYTDYIT-ADVVDRIITAINVAGTTPVGL-NIAQ 118
 Qy 102 DISLGSVVDTCGANKLAVNIDGLTLTGTAAGANPALLFGGQAAANNVYALGN 161
 Db 119 NTVGSIITRG--NLLPVTI-AGKSLILNGTNAVA--AN-----HGPDPADN-YTGLGN 167
 Qy 162 ITLGGANAGLTASDPVGLGPITLAGNIDGGIITDNTDAINGT 206
 Db 168 ITLGGANAELIQSATP--AKITLAGNIDGGIITVTKDAAINGT 210

RESULT 10

Q9KKB1 PRELIMINARY; PRT; 1618 AA.
 AC Q9KKB1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE Outer membrane protein A (Fragment).
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 gene coding the outer-membrane protein ompB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 DR EMBL; AF123713; AAF74116.1;
 DR InterPro; IPR003858; rCmpA_rompB.
 DR Pfam; PF02708; rCmpA_rompB; 1.
 FT NON_TER 1618
 SQ SEQUENCE 1618 AA; 164193 MW; 66AD2199620750C6 CRC64;

Query Match 12.18; Score 366.5; DB 2; Length 1618;
 Best Local Similarity 24.18; Pred. No. 8.4e-07;

Matches 179; Conservative 91; Mismatches 251; Indels 223; Gaps 37;
 Qy 23 TTSTAAILMLTSGS---VLGAARTVTDAGELAA-----GTNIGPGAGAFVAGSTLTQYTGAF 75
 Db 10 TTNGVATTVDGSGFDQTVLANVAVAPNAVITANANGINLNTPAGSF-----NGLF 61
 Qy 76 TVTDADVSRALDLNPPAAGLFSVTGDISLGSVVD-TGGANKLAVNIDGLTLTLTG- 133
 Db 62 -----LSNANLAV---TVSEDTTLGFNNANANRNFNLTDAGTKTLITGQGI 108
 Qy 134 ----TAAYGA-PPALLFGGQAAANNVYALGNITLGGANAGLT---IASDPVGLGPITL 185
 Db 109 TNVOSAATHNAQNIIVAKENGGAIAANDLSGLGTIDFGAAASTLVFLANPTTQKAPLIL 168
 Qy 186 AGNIDGGIITDNTDAAINGTIGNTPAAQIS-----IG----- 219
 Db 169 ADN---ALIVNGA---NGTLNVNTNGFIQVSDKSPATVKAINIGDGGQGFMTNATNAN 220
 Qy 220 ASTLSLGGAVIKATTTKLTAAPVLTITNANAVLTG-ADVNTTGGDDVGVNLNGAL--S 276
 Db 221 ALNLQAGGTTINFNGTDGTGR---LVLLSKNGAATDFNVTGSLGNNLKGIIELNTVAING 277
 Qy 277 QVTGNIGNTSLATISVGAG-----TATLGG-----AVIKAT----- 308
 Db 278 QLIANAPANAVIGTNNAGRAAGFVVDNGKRAATIDGQVYAKDMVIOSSANANGQVNER 337
 Qy 309 -----TTKLTAASVLTITNNAVLTGAVD-----NTTG-----G 336
 Db 338 HIVDVGIDGTTAFKTAASIVAITONSFGTTDFGNLAAQVTPDPTMTLTGNTGDTANNPG 397
 Qy 337 DNVGVNLSG---ALSOVTCNIGNTSLATINI-GAGVATLDGAVIKATTKLTDDASV 391
 Db 398 NTAGVITTAANGTLASASADANNAVITNITAEASGVGVVQLSGT---HTAELRLGNAGSV 455
 Qy 392 LIETNPVVVTGAIDNT---GNANKGVVIFTGASTVTDNIGN---TAVLAESV----- 438
 Db 456 FKLADGVINGKVNQTVLVGVLAAGAITLDGSATITGDIINGGGGAALOSITLANDATK 515
 Qy 439 -----GAGLIQIQGVVKANA-----INLTD-----NASVVTFTG 468
 Db 516 TLTLLGANAIIISANGTTFNFQANGTKIKLTSTONNVVDCDLAIATDGTGVVDASLTLNAQ 575
 Qy 469 DSTVTGSGIS---GTFLFATVNIAGITLRAGGSLAANNIDFGAASNLEF----- 514
 Db 576 TLTISGTIGIIGANTTLGQFNIGSSKTYTLNGGNVAINELVIGNNGSVQFAHNTYLTIRT 635
 Qy 515 NGPAGKNYNLIGTIANGNATNLNAAAGTVIANDVSTGTVQAQINQNNKIFVINAKNAV 574
 Db 636 TNAAGQCKIIFNPVNV-NNTTL---AAGTNLGS---AANPLAEIN-----FGSKGARA 681
 Qy 575 D-IILDAQAISFKGAASRLFLANVS 597
 Db 682 DTVLNV-----GEGVNLVATNIT 699

RESULT 11

Q93QW9 PRELIMINARY; PRT; 1654 AA.
 AC Q93QW9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE Outer membrane protein A (Fragment).
 OS Rickettsia felis (Rickettsia azadi).
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=42862;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21183306; PubMed=11289671;
 RA Moron C.G., Bouyer D.H., Yu X.J., Foil L.D., Crocquet-Valdes P.,
 Walker D.H.;

RESULT 13
Q9KKA5
ID Q9KK

[illegible]

```

Db 662 SARNLSLNTGTGSDVNTNGTVSARAAATIA-----AGNLVNRGGTLVAVGDVTAH 711
Qy 401 TGAIDNTGNA---NKGVVIFTGASTVTDNIGNTAVLA-EVSVGAGLLQIQGVVKANAIN 456
Db 712 VGSLDNTGGAGFGSGGLNVTSGAI-DNAGGKLVAAODATLNAASLGNQGGTISARNLS 770
Qy 457 LTDNASVVTFGTGSDVNTGSGITGELFATVNIAGITLRAGGSAA-----NNID 505
Db 771 LN-----TGTGAIDNTGTVSAGTATVDAGSLINQGGTLVAVADVQAVNGRLDNTG 822
Qy 506 FGAASNFENGPAGKNNYLIETIANGNN---ATLINAAGTVIANDVSGITVQAQINIQNN 562
Db 823 GGLGSGQ-----SGGLNVTSGAIDNAGKLVAAODASLTGTSLGNOA-GTVAGRNV--- 872
Qy 563 KIFVINAKNADVDIILD-----AQAISEKGAASRLFLANVSLQ 599
Db 873 -----TVNTGTGALDNTGGTAIAAAGLDATAGALTANANGVMQ 909

RESULT 15
Q9KKA8
ID Q9KKA8 PRELIMINARY; PRT: 1615 AA.
AC Q9KKA8:
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE OmpB (Fragment).
CN OmpB.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M/5-6;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL; AF123716; AAF34119.1;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1615
SQ SEQUENCE 1615 AA; 162788 MW; 932812728425FB73 CRC64;

Query Match 11.9%; Score 361; DB 2; Length 1615;
Best local Similarity 25.3%; Pred. No. 1.4e-06;
Matches 188; Conservative 87; Mismatches 249; Indels 218; Gaps 40;

Qy 23 TTSTAAILTSGVLGAARTVTADGAELAA-----GTNIGPGAGAFVAGSTLOYTGAF 75
Db 10 TTNGRATTVDGAGFDQTAQVNAAPAVITANANGINFNTPAGSF-----NGLF 61
Qy 76 TVTDADVSRALDLNNFAAGLSVTDISLSGVVD-TGGANKLAVNIDDLTLTLTGTO- 133
Db 62 LCT-----ANNLAV---TVSADTLTGFVTNVNNGSNFLTLGAGKTLTITGQGI 108
Qy 134 TAAYGA-----NPALLFGGGAANNVTYALGNITLGCANAGL--TIASDDVVLGPITL 185
Db 109 TNAQAAVTNNAAQVNGVAVANNDLSSGCAIDFGAAASTLVFNLAEPPTQKAPLIL 168
Qy 186 AGN-----IDGGGIITDN-----TDAAIN-----GTIGNTPAAQISIGASTLS 224
Db 169 GGNVAVANGVNTLVNTNGFIQVSDKSFATVKAINIGDQGGFI FNTNVAA-----GGNALN 224
Qy 225 L--GGAVIKATTKLTNAAPVLTITNANAVLTG-AVDNTTGGDDVGLNLN--GALSQVT 279
Db 225 LQVGGATINFNGTGDGTGR---LVLLSKNGAATDFNVNLSLGNLKGIIIEFNTVAVAGQLI 281
Qy 280 GNIGNTNSLA-----TISVGAG-TATLGG-----AVIKAT----- 308
Db 282 ANAGPANAVIGTNDAGRAAGFWVSVGNGNAATITGVYAKDMVITQSANAGGVNFGHIV 341

```

```

Qy 309 -----TTKLTAASVLTLTNAVLTGAVD-----NTTGG-----DNV 339
Db 342 DVGTDGTTAFATAASKVAITONSFGTTDFGNLAAQIIVPDTMTLTGNTGNASNPETA 401
Qy 340 GVVNLGSG-----ALSOVTGNIGNTNSLATINI-GAGVATLDGAVIKATTTKLTDASVLIF 394
Db 402 GVIITFAANGTLASASADANVAVTNNITAEASGVGVQVLSGT--HTAELRLGNAGSVFKL 459
Qy 395 TNPVVVTGAIDNT-----GNANKGVVIFTGASTVTDNIGN--TAVLAESV----- 438
Db 460 ADGTVINGKVNQTA VVGALAAAGAITLDGSATITGIDGNGGAAALOGITLANDATKLT 519
Qy 439 --GAGLL-----QIQGGVVKAN-----AINLTD-----NASVWTFGTGDS 470
Db 520 LGGANIIGANGGTINFGANGGTIKLTSTONNILLVDFDLAIN-TDQTGVVDASSLINAQTL 578
Qy 471 TVTGSIG-----GTFLFATVNIAGITLRAGGSLAANNIDFGAASNLEF-----NG 516
Db 579 TISGTIGTVGANNTKLGQFNVGSSKLTALNGGNVAINELVIGNNGSVQFAHDTYLTITKTTN 638
Qy 517 PAGKNYNLIGTIANGNNATLINNAAGTVIANDVSGITVQAQINIQNNKIFVINAKNADVD- 575
Db 639 AAGQGIIFNPVNVN-NNTTL---AAGTNLGS--AANPLAEIN-----FGSKGAHADT 684
Qy 576 ILDAQAISFKGAASRLFLANVS 597
Db 685 ILNV-----GKGVNLYATNIT 700

```

Search completed: April 14, 2003, 16:11:05
Job time : 102 secs

THIS PAGE BLANK (USPTO)